(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/36674 A2

(51) International Patent Classification[7]: C12Q 1/68, G01N 33/53, A61P 43/00

(21) International Application Number: PCT/GB00/04267

(22) International Filing Date: 8 November 2000 (08.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9926805.4 13 November 1999 (13.11.1999) GB

(71) Applicant *(for all designated States except MG, US)*: ASTRAZENECA AB [SE/SE]; S-151 85 Sodertalje (SE).

(71) Applicant *(for MG only)*: ASTRAZENECA UK LIMITED [GB/GB]; 15 Stanhope Gate, London W1Y 6LN (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: BULL, John, Henry [GB/GB]; 25 Pownall Street, Macclesfield, Cheshire SK10 1DF (GB). ELLISON, Gillian [GB/GB]; Alderley Park, Macclesfield, Cheshire SK10 4TG (GB). PASKINS, Lynn, Dora [IE/GB]; 45 St. Anne's Avenue, Grappenhall, Warrington WA4 2PL (GB).

(74) Agent: GAINEY, Laurence, David, Scott; Astrazeneca, Global Intellectual Property, P.O. Box 272, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4GR (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/36674 A2

(54) Title: DIAGNOSTIC METHODS

(57) Abstract: The present invention derives from the identification of 26 gene transcripts (Markers) that exhibit aberrant expression levels in prostate disorder tissues. The invention therefore relates to diagnostic techniques for the detection of human prostate disorders, such as cancer, by detecting one or more of these Markers, intermediates, precursors or products (mRNA, cDNA, genomic DNA, or protein). The invention is also directed to methods for identifying modulators of prostate disorders, which modulators, such as chemical compounds, antisense molecules and antibodies interact with and modulate any one of the Markers identified.

RTS-0303

In early disease, where the cancer is confined to the prostate gland and has not spread beyond the gland, the tumour can potentially be removed by surgery. Surgery is a common treatment of prostate cancer. Transurethral resection is a procedure in which the cancer is cut from the prostate via the urethra. A radical prostatectomy is where the whole prostate and some of the tissue around the gland is removed. Radiation therapy is also used to treat early prostate cancer.

Once disease has become metastatic, its course is rapid and relatively predictable, leading to death within 2-5 years if untreated.

Prostate cancer commonly occurs in men above the age of 50 years. The incidence of prostate cancer rises sharply between the ages of 60 and 80 years, with more than 80% of all cases of the diseases being diagnosed in men over 65 and less than 1% in men under 50 years of age. The incidence and prevalence of prostate cancer have increased considerably throughout the world over the past two decades. The average age of detection is 72 in the UK and 66 in the US. Active screening programmes may lead to a lowering of the average age at first diagnosis.

Routine medical examinations for obstructive urinary symptoms may lead to an initial diagnosis of prostate cancer. When prostate cancer is asymptomatic, there are two tests that can be performed to detect the presence of cancer: digital rectal examination (DRE) and prostate specific antigen (PSA). PSA is an enzyme which is secreted almost exclusively by epithelial cells in the prostate; measurement of the quantity of PSA in the blood provides a more reliable indicator of the presence of prostate cancer than the DRE test. However, elevations in serum levels of PSA can also be found in patients with benign prostatic hyperplasia (BPH) and prostatitis. In some cases, where a patient has elevated levels of PSA in his blood, a third diagnostic technique is used called 'transrectal ultrasonograpy' (TRUS), which can further substantiate the presence of prostate cancer. In conjunction with taking a biopsy of the tissue, TRUS has been shown to increase the detection rate of prostate cancer two-fold, when compared to DRE alone. The combination of digital rectal examination, serum PSA level and transrectal ultrasound is currently the best available diagnostic tool (Gorgoulis VG. et al. Anticancer Res 19:2327-2348 (1999)).

It is common for the prostate gland to become enlarged with ageing, a condition called benign prostatic hyperplasia (BPH), or benign prostatic hypertrophy. Severe BPH can cause

nucleic acid targeted to the Marker mRNA; or more widely in the identification or development of chemical or hormonal therapeutic agents. The person skilled in the art is also capable of devising screening assays to identify compounds (chemical or biological) that modulate (activate or inhibit) one or other of the identified Marker, which compounds may prove useful as therapeutic agents in treating a prostate disorder. Monitoring the Markers could also be useful in identifying inhibitors, antagonists or biochemical signalling effects in high throughput screening.

According to a first aspect of the invention there is provided a method for diagnosing or prognosing or monitoring a prostate disorder comprising testing a biological sample for aberrant levels of one or more of the Markers selected from the group consisting of Marker 1 to Marker 26. Preferred Markers are numbers: 8, 13, 14, 15, 17, 18, 19, 20, 22, 24, 25 and 26.

The invention lies in the identification of these differentially expressed Markers in prostate disorders. Accordingly, the invention is directed to any diagnostic method capable of assessing the differential expression levels, relative to expression in control tissues, of one or more of the identified Markers, either alone or as a panel. In particular, such methods include assessment of mRNA transcript levels and/or protein levels. The presence of aberrant expression levels of one of the Markers indicating the presence of a prostate disorder.

In a preferred embodiment the diagnostic method involves testing for more than one of the Markers identified herein. As separate independent embodiments the diagnostic method may involve testing for 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or all of the Markers identified herein, optionally as part of a panel test of other gene Markers.

It will be apparent to the person skilled in the art that there are a large number of analytical procedures which may be used to detect the amount of any of the Marker products present in a test sample.

The test sample comprising nucleic acid or protein is conveniently any prostatic material or biological sample, including TURP chip, biopsy, excised prostate or part thereof, a sample of bone marrow aspirate, bone marrow biopsy, lymph node aspirate, lymph node biopsy, spleen tissue, fine needle aspirate, skin biopsy or organ tissue biopsy, particularly prostate gland tissue, invaded surgical margin, invaded lymph node, invaded lung, invaded

biomarkers for distinguishing cancer from BPH. Highly specific yet low abundance Markers could be invaluable if secreted in protein form, as they could permit non-invasive testing of body fluids.

According to a further aspect of the invention there is provided a method for distinguishing prostate cancer from BPH comprising testing a biological sample for aberrant levels of one or more of the Markers selected from the group consisting of Marker 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, 18, 20, 23, 25 and 26. Such a method might involve measuring the expression levels of one or more of these Markers in a biological sample taken from an individual, comparing the Marker expression level detected with control values or historical values and depending on the value detected determining whether or not the individual has prostate cancer or BPH.

Each of the Marker genes useful in the present invention are already in the public domain. However, the inventors are not aware of any prior art disclosing an association of any of the Marker genes with prostate cancer. Thus, although each of the Markers/genes identified herein are themselves already known, their association with prostate disorders leading to the present invention is unknown.

Marker 1 is the human smooth muscle protein, 22kDa. The gene coding for this protein has been cloned and sequenced (Thweatt et al., Biochem. Biophys. Res. Commun. 187:1-7 (1992)). These authors identify 22kD smooth muscle protein as a fibrobast or smooth muscle protein. There is no teaching in this paper of an association with prostate cancer. The sequence of 22kDa smooth muscle protein is disclosed in Thweatt et al. and is present in the EMBL database under accession number HS22SM. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HS22SM, unless stated otherwise or apparent from the context.

Marker 2 is the Rho GDP dissociation inhibitor protein. The gene coding for this protein has been cloned and sequenced (Leffers et al., Exp. Cell Res. 209:165-174 (1993)). These authors identify Rho GDP dissociation inhibitor protein as a keratinocyte protein. There is no teaching in this paper of an association with prostate cancer. The sequence of Rho GDP dissociation inhibitor protein is disclosed in Leffers et al., and is present in the EMBL database under accession number HSRHO1. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL

acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSD966, unless stated otherwise or apparent from the context.

Marker 7 is the human mRNA KIAA0120. This cDNA has been cloned and sequenced (Nagase et al., DNA Res. 2:37-43 (1995)). These authors identify KIAA0120 as a brain transcript. There is no teaching in this paper of an association with prostate cancer. The sequence KIAA0120 is disclosed in Nagase et al., and is present in the EMBL database under accession number HSORFF. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSORFF, unless stated otherwise or apparent from the context.

Marker 8 is the human ribosomal protein S25. The gene coding for this protein has been cloned and sequenced (Li et al., Gene 107:329-333 (1991)). These authors identify ribosomal protein S25 as a leukaemia cell line protein. There is no teaching in this paper of an association with prostate cancer. The sequence of human ribosomal protein S25 is disclosed in Li et al., and is present in the EMBL database under accession number HSRPS25. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSRPS25, unless stated otherwise or apparent from the context.

Marker 9 is the human 80K-H protein (kinase C substrate). The gene coding for this protein has been cloned and sequenced (Sakai et al., Genomics 5:309-315 (1989)). These authors identify 80K-H protein (kinase C substrate) as a fibroblast and epidermal carcinoma cell protein. There is no teaching in this paper of an association with prostate cancer. The sequence of 80K-H protein (kinase C substrate) is disclosed in Sakai et al., and is present in the EMBL database under accession number HSG19P1A. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSG19P1A, unless stated otherwise or apparent from the context.

Marker 10 is the human alpha-2-macroglobulin protease inhibitor protein. The gene coding for this protein has been cloned and sequenced (Kan et al., Proc. Natl. Acad. Sci. U.S.A. 82:2282-2286 (1985)). These authors identify alpha-2-macroglobulin protease inhibitor protein as a serum protein. There is no teaching in this paper of an association with prostate cancer. The sequence of alpha-2-macroglobulin protease inhibitor protein is disclosed in Kan et al., and is present in the EMBL database under accession number

referred to herein, refers to that in the EMBL database under accession number HSLB2A26, unless stated otherwise or apparent from the context.

Marker 15 is the human PRSM1 protein. The gene for PRSM1 has been cloned and sequenced (Scott et al., Gene 174:135-143 (1996)). These authors identify PRSM1 as a putative metallopeptidase. There is no teaching in this paper of any association with prostate cancer. The sequence of PRSM1 is disclosed in Scott et al. and is present in the EMBL database under accession number HSU58048. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSU58048, unless stated otherwise or apparent from the context.

Marker 16 is the human ribosomal large subunit L12 protein. The gene for ribosomal large subunit L12 protein has been cloned and sequenced (Chu et al., Nucleic Acids Res. 21:749-749 (1993)). These authors identify human ribosomal large subunit L12 as a ribosomal protein. There is no teaching in this paper of any association with prostate cancer. The sequence of human ribosomal large subunit L12 is disclosed in Chu et al. and is present in the EMBL database under accession number HSL12A. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSL12A, unless stated otherwise or apparent from the context.

Marker 17 is the human 7SK45 RNA. The gene for 7SK45 RNA has been cloned and sequenced (Murphy et al., Nucleic Acids Res. 14:9243-9260 (1986)). These authors identify human 7SK45 as a small cytoplasmic RNA. There is no teaching in this paper of any association with prostate cancer. The sequence of 7SK45 is disclosed in Murphy et al. and is present in the EMBL database under accession number HS7K45. For the purpose of this application, the gene sequences and sequence positions referred to herein, refers to that in the EMBL database under accession numbers: HS7K45, unless stated otherwise or apparent from the context.

Marker 18 is the human KIAA0588 protein (human protocadherin gamma). This protein is abundantly expressed in brain cells. The gene for this protein has been cloned and sequenced (Nagase et al., DNA Research. 5:31-39 (1998)). There is no teaching in this paper of an association with prostate cancer. The sequence of the KIAA0588 protein is present in the EMBL database under accession number AB011160. For the purpose of this application,

application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSSRP14A, unless stated otherwise or apparent from the context.

Marker 23 is bone small proteoglycan I (biglycan) protein. The gene for this protein has been cloned and sequenced (Fisher et al., J. Biol. Chem. 264:4571-4576 (1989)). These authors identify bone small proteoglycan I (biglycan) as a protein derived from human bone-derived cells. There is no teaching in this paper of an association with prostate cancer. The sequence of bone small proteoglycan I (biglycan) protein is disclosed in Fisher et al. and is present in the EMBL database under accession number HSHPGI. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSHPGI, unless stated otherwise or apparent from the context.

Marker 24 is the human KIAA0045 gene. The gene for this protein has been cloned and sequenced (Nomura N et al., DNA Res 1:223-229 (1994)). These authors identify KIAA0045 as a protein from the human immature myeloid cell line KG-1. There is no teaching in this paper of an association with prostate cancer. The sequence of KIAA0045 is disclosed in the EMBL database under accession number HSKG1C (accession no. D28476). For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number HSKG1C, unless stated otherwise or apparent from the context.

Marker 25 is b-cell receptor associated protein, also known as REA (repressor of estrogen activity). The gene for this protein has been cloned and sequenced (Montano et al., Proc Natl Acad Sci U.S.A. 96:6947-6952 (1999)). These authors identify REA as a protein derived from human breast. There is no teaching in this paper of an association with prostate cancer. The sequence of b-cell receptor associated /REA protein, is disclosed in Montano et al. and is present in the EMBL database under accession number AF150962. For the purpose of this application, the gene and amino acid sequences and sequence positions referred to herein, refers to that in the EMBL database under accession number AF150962, unless stated otherwise or apparent from the context.

Marker 26 is the cystatin B protein. The gene for cystatin B has been cloned and sequenced (Pennacchio et al., Science 271:1731-1734 (1996)). These authors identify cystatin B protein as associated with progressive myoclonus epilepsy disease. There is no teaching in

markers such as PSA, prostatic acid phosphatase, prostate-specific membrane antigen or others (e.g. those reviewed in Gao et al., Prostate 31:264-281 (1977)); markers associated with cell proliferation, invasiveness, tumour angiogenesis, metastasis or any other aspect of carcinogenesis; or markers for other diseases and conditions. Levels of Marker mRNA in the test sample can be detected by any technique known in the art. These include Northern blot analysis, reverse transcriptase-PCR amplification (RT-PCR), microarray analysis and RNAse protection.

In one embodiment, levels of Marker RNA in a sample can be measured in a Northern blot assay. Here, tissue RNA is fractionated by electrophoresis, fixed to a solid membrane support, such as nitrocellulose or nylon, and hybridised to a probe or probes capable of selectively hybridising with the Marker RNA to be detected. The actual levels may be quantitated by reference to one or more control housekeeping genes. Probes may be used singly or in combination. This may also provide information on the size of mRNA detected by the probe. Housekeeping genes are genes which are involved in the general metabolism or maintenance of the cell, and are considered to be expressed at a constant level irrespective of cell type, physiological state or stage in the cell cycle. Examples of suitable housekeeping genes are: beta actin, GAPDH, histone H3.3 or ribosomal protein L13 (Koehler et al., Quantitation of mRNA by Polymerase Chain Reaction. Springer-Verlag, Germany (1995)).

To gauge relative expression levels, a control sample can be run alongside the test sample or, the test result/value can be compared to Marker expression levels expected in a normal or control tissue. These control values can be generated from prior test experiments using normal or control tissues, to generate mean or normal range values for each Marker.

In another embodiment, the Marker nucleic acid in a tissue sample is amplified and quantitatively assayed. The polymerase chain reaction (PCR) procedure can be used to amplify specific nucleic acid sequences through a series of iterative steps including denaturation, annealing of oligonucleotide primers (designed according to the published Marker sequence to be detected), and extension of the primers with DNA polymerase (see, for example, Mullis, et al., U.S. patent No. 4,683,202; Loh et al., Science 243:217 (1988)). In reverse transcriptase-PCR (RT-PCR) this procedure is preceded by a reverse transcription step to allow a large amplification of the number of copies of mRNA (Koehler et al., *supra*). Other known nucleic acid amplification procedures include transcription-based amplification systems (TAS) such as nucleic acid based sequence application (NASBA) and 3SR (Kwoh et

A variety of methods are currently available for making arrays of biological molecules. The 'dot or slot blot' approach, whereby an ordered array of DNA is vacuum blotted using a manifold, or hand blotted by capillary action, onto a porous membrane, such as nylon or nitrocellulose has been around for many years (Maniatis et al., Molecular Cloning-A Laboratory Manual, First Edition, Cold Spring Harbor, (1982)). Methods for preparing a plurality of oligonucleotide sequences and for attaching these to solid supports at high density are also known in the art. For example, US Patent No. 4,562,157 describes a method of using photo-activatable cross-linking groups to immobilise pre-synthesised ligands on surfaces. Fodor et al. (Nature 364:555-556 (1993)) and US Patent No. 5,143,854 describe the 'light-directed chemical synthesis' method for synthesising ligands, including oligonucleotides, directly onto a substrate surface at the desired location. US 5,700,637 also describes methods for *in situ* synthesis of oligonucleotides on solid support surfaces. In addition, such methods for preparing microarrays can easily be automated. International Publication No. WO 95/35505 discloses an automated capillary dispensing device and method for applying biological macromolecules to solid supports. International Publication No. WO 97/44134 also describes devices for delivery of small volumes of liquid (which may contain biological macromolecules) in a precise manner to produce microsized spots on a solid surface to generate a microarray. Similarly, International Publication No. WO 98/10858 also describes an apparatus for the automated synthesis of molecular arrays. Techniques exist for applying the oligonucleotides to the array at high density and for example, techniques exist for applying well in excess of 10^3 distinct polynucleotides per 1 cm^2.

Microarray technology makes it possible to simultaneously study the expression of many thousands of genes in a single experiment. Analysis of gene expression in human tissue (e.g. biopsy tissue) can assist in the diagnosis and prognosis of disease and the evaluation of risk for disease. A comparison of levels of expression of various genes from patients with defined pathological disease conditions with normal patients enables an expression profile, characteristic of disease, to be created. There are currently two main approaches to analyse gene expression using microarrays. In the first approach, cDNA fragments, often generated by PCR, for each of the genes under study are attached to an array. Typically, mRNA isolated from the test samples (i.e. induced or un-induced) is reverse transcribed into cDNA with incorporation of a fluorescent label. The cDNA is sheared and hybridised to the array. If a control test sample is to be run at the same time, mRNA from this sample can be reverse

In transcript profiling, several or many mRNAs are detected in the same procedure. One or more of these mRNAs may be diagnostic of cancer cells (i.e prostate cancer) in a tissue sample. In one embodiment, combinations of probes can be used to classify the cancer cells into clinically relevant types, according to the complex expression pattern of Markers measured on the array. Such classification may help to define which tumours are growing aggressively, or harbour latent signs of aggression, or are less aggressive or benign. The array provides a quantitative measure of Marker RNAs. This is done by comparison of Marker RNA signal with control signal. In a preferred embodiment hybridisation signals generated are measured by computer software analysis of images on phosphorimage screens exposed to radioactively labelled tissue RNA hybridised to a microarray of probes on a solid support such as a nylon membrane. In another, quantities are measured by densitometry measurements of radiation-sensitive film (e.g. X-ray film), or estimated by visual means. In another embodiment quantities are measured by use of fluorescently labelled probe, which may be a mixture of tumour and normal RNA differentially labelled with different fluorophores, allowing quantities of Marker mRNAs to be expressed as a ratio versus the normal level. The solid support in this type of experiment is generally a glass microscope slide, and detection is by fluorescence microscopy and computer imaging.

The detection of specific interactions may be performed by detecting the positions where the labelled target sequences are attached to the array. Radiolabelled probes can be detected using conventional autoradiography techniques. Use of scanning autoradiography with a digitised scanner and suitable software for analysing the results is preferred. Where the label is a fluorescent label, the apparatus described, e.g. in International Publication No. WO 90/15070, US Patent No. 5, 143,854 or US Patent No. 5,744,305 may be advantageously applied. Indeed, most array formats use fluorescent readouts to detect labelled capture:target duplex formation. Laser confocal fluorescence microscopy is another technique routinely in use (M.J.Kozal et al., Nature Medicine 2:753-759 (1996)). Mass spectrometry may also be used to detect oligonucleotides bound to a DNA array (Little et al, Analytical Chemistry 69: 4540-4546, (1997)). Whatever the reporter system used, sophisticated gadgetry and software may be required in order to interpret large numbers of readouts into meaningful data (such as described, for example, in US Patent No. 5,800,992 or International Publication No. WO 90/04652).

RNA can be extracted at a later date. Proprietary reagents are available which allow tissue or cells to be conveniently stored for several days at room temperature and up to several months at 4°C (e.g. RNAlater, Ambion Inc., TX). Prior to extraction, methods such as grinding, blending or homogenisation are used to dissipate the tissue in a suitable extraction buffer. Typical protocols then use solvent extraction and selective precipitation techniques. Example 1 describes such a method.

In another embodiment, tissue is directly analysed for the presence of Marker nucleic acid. This can be by *in situ* hybridisation, where sections of tissue may be interrogated with specific probes to determine which morphological cell type in the sample displays a marker nucleic acid, such as sequences corresponding to Markers 1-26. *In situ* hybridisation typically comprises 3 steps. Firstly tissue is fixed, and sections are prepared by standard treatments known to those in the art (Polack and McGee, *In situ* hybridisation: principles and practice. Oxford University Press, 1998). Secondly, Marker mRNA and amplified Marker DNA can be detected by hybridisation with e.g. a biotin-, digoxygenin- or radio-labelled Marker probe, typically for 2-16 hours at 42°C in a suitable hybridisation buffer. A typical buffer might contain 50% formamide, 5% dextran sulphate, 2xSSC and 10-20ng of probe per 7µl (Herrington and McGee, Diagnostic Molecular Pathology. IRL Press, Oxford, 1992). The probe can be made of DNA or RNA. Lastly, following stringency washes, the probes in hybridisation complexes are detected with chromogenic or fluorescent reagents, which can be visualised by microscopy, or by autoradiography in the case of radiolabelled probes. Signal amplification systems using e.g. tyramide can be used to increase sensitivity (Polack and McGee, *supra)*. *In situ* PCR using oligonucleotide probes complementary to the nucleic acid of any of Markers 1 - 26 is therefore envisaged. *In situ* hybridisation can follow *in situ* PCR, giving greater specificity (Polack and McGee, *supra*). Techniques for quantitation of signal, and quantification of positive cells in a section are available to the pathologist using image analysis. *In situ* Marker visualisation permits localisation of signal in mixed-tissue specimens commonly found in tumours, and is compatible with many histological staining procedures. In one embodiment, several probes can be differentially labelled and hybridised simultaneously to the same section, and detected using appropriate reagents. In another, serial sections from the same sample can be analysed with a panel of probes. Quantitation may involve comparison with one or more control housekeeping genes as discussed above.

cDNA) then the probe or primer sequence can hybridise to the sense or antisense strand. If however the target is mRNA (single stranded sense strand) the primer/probe sequence will have to be the antisense complement.

An example of a suitable hybridisation solution when a nucleic acid is immobilised on a nylon membrane and the probe nucleic acid is greater than 500 bases or base pairs is: 6 x SSC (saline sodium citrate), 0.5% SDS (sodium dodecyl sulphate), 100µg/ml denatured, sonicated salmon sperm DNA. The hybridisation being performed at 68°C for at least 1 hour and the filters then washed at 68°C in 1 x SSC, or for higher stringency, 0.1 x SSC/0.1% SDS.

An example of a suitable hybridisation solution when a nucleic acid is immobilised on a nylon membrane and the probe is an oligonucleotide of between 12 and 50 bases is: 3M trimethylammonium chloride (TMACl), 0.01M sodium phosphate (pH 6.8), 1mM EDTA (pH 7.6) , 0.5% SDS,100µg/ml denatured, sonicated salmon sperm DNA and 0.1 dried skimmed milk. The optimal hybridisation temperature (Tm) is usually chosen to be 5°C below the Ti of the hybrid chain. Ti is the irreversible melting temperature of the hybrid formed between the probe and its target. If there are any mismatches between the probe and the target, the Tm will be lower. As a general guide, the recommended hybridisation temperature for 17-mers in 3M TMACl is 48-50°C; for 19-mers,it is 55-57°C; and for 20-mers, it is 58-66°C.

A suitable hybridisation protocol is described in Example 5, however, operable variations to this method will be apparent to the person skilled in the art.

Where the cDNA molecules of Markers 1 - 26 encode proteins or parts of cellular proteins, these may themselves act as prostate disease Markers. To detect proteins in tissue, cells, body fluids or extracts of these sample types, specific antibody can be used. These antibodies can be prepared using the Marker protein/polypeptides.

Methods of making and detecting labelled antibodies are well known (Campbell; Monoclonal Antibody Technology, in: Laboratory Techniques in Biochemistry and Molecular Biology, Volume 13. Eds: Burdon R et al. Elsevier, Amsterdam (1984)). The term antibody includes both monoclonal antibodies, which are a substantially homogeneous population, and polyclonal antibodies which are heterogeneous populations. The term also includes inter alia, humanised and chimeric antibodies. Monoclonal antibodies to specific antigens may be obtained by methods known to those skilled in the art, such as from hybridoma cells, phage display libraries or other methods. Monoclonal antibodies may be inter alia, human, rat or

that the nucleic acid sequence encoding the polypeptide is transcribed into RNA in the host cell transformed or transfected by the expression vector construct. The coding sequence may or may not contain a signal peptide or leader sequence for secretion of the polypeptide out of the host cell. Expression and purification of the Marker polypeptides can be easily performed using methods well known in the art (for example as described in Sambrook et al. *supra*).

The Marker polypeptides so produced can then be used to inoculate animals, from which serum samples, containing the specific antibody against the introduced Marker protein/polypeptide, can later be obtained.

Rodent antibodies may be humanised using recombinant DNA technology according to techniques known in the art. Alternatively, chimeric antibodies, single chain antibodies, Fab fragments may also be developed against the polypeptides of the invention (Huse et al., Science 256:1275-1281 (1989)), using skills known in the art. Antibodies so produced have a number of uses which will be evident to the molecular biologist or immunologist skilled in the art. Such uses include, but are not limited to, monitoring enzyme expression, development of assays to measure enzyme activity and use as a therapeutic agent. Enzyme linked immunosorbant assays (ELISAs) are well known in the art and would be particularly suitable for detecting the Marker proteins or polypeptide fragments thereof in a test sample.

The expression system described in Example 3 can produce protein for use as an antigen for the generation of antibodies for use in an ELISA assay to detect Marker protein in body fluids or by immunohistochemistry (as described in Example 4) or other means. In addition, an antibody could be used individually or as part of a panel of antibodies, together with a control antibody which reacts to a common protein, on a dipstick or similar diagnostic device.

Levels of Marker gene expression can also be detected by screening for levels of polypeptide (Marker protein). For example, monoclonal antibodies immunoreactive with a Marker protein can be used to screen a test sample. Such immunological assays can be done in any convenient format known in the art. These include Western blots, immunohistochemical assays and ELISA assays. Functional assays can also be used, such as protein binding determinations.

In another preferred embodiment antibodies directed against a Marker protein or proteins can be used, to detect, prognose, diagnose and stage prostate cancer or its precursor lesions, or related prostate disorders. Various histological staining methods known in the art,

Thus, according to another aspect of the invention there is provided a method for treating a patient suffering from a prostate disorder comprising administering to said patient an effective amount of an anti-sense molecule capable of binding to the mRNA of a Marker gene selected from the group consisting of: Marker 1 - Marker 26, and inhibiting expression of the protein product of the Marker gene.

Complete inhibition of protein production is not essential, indeed may be detrimental. It is likely that inhibition to a state similar to that in normal tissues would be desired.

This aspect of antisense therapy is particularly applicable if the prostate disorder is a direct cause of over-expression of the Marker gene(s) in question, although it is equally applicable if said Marker gene(s) indirectly cause the prostate disorder. Having identified the particular Marker genes (1 - 26) over-expressed in prostate disorders, and with knowledge of the gene and mRNA sequence the person skilled in the art is able to design suitable antisense nucleic acid therapeutic molecules and administer them as required.

Antisense oligonucleotide molecules with therapeutic potential can be determined experimentally using well established techniques. To enable methods of down-regulating expression of a Marker gene of the present invention in mammalian cells, an example antisense expression construct can be readily constructed for instance using the pREP10 vector (Invitrogen Corporation). Transcripts are expected to inhibit translation of the gene in cells transfected with this type of construct. Antisense transcripts are effective for inhibiting translation of the native gene transcript, and capable of inducing the effects (e.g., regulation of tissue physiology) herein described. Oligonucleotides which are complementary to and hybridisable with any portion of Marker gene mRNA are contemplated for therapeutic use. U.S. Patent No. 5,639,595, "Identification of Novel Drugs and Reagents", issued Jun. 17, 1997, wherein methods of identifying oligonucleotide sequences that display *in vivo* activity are thoroughly described, is herein incorporated by reference. Expression vectors containing random oligonucleotide sequences derived from Marker polynucleotides are transformed into cells. The cells are then assayed for a phenotype resulting from the desired activity of the oligonucleotide. Once cells with the desired phenotype have been identified, the sequence of the oligonucleotide having the desired activity can be identified. Identification may be accomplished by recovering the vector or by polymerase chain reaction (PCR) amplification and sequencing the region containing the inserted nucleic acid material. Antisense molecules can be synthesised for antisense therapy. These antisense molecules may be DNA, stable

analogous to quantitative RT-PCR already described. Marker gene amplification could also be detected *in situ* in cells by probe hybridisation. Fluorescently-labelled probes (used in Fluorescent in situ Hybridisation or FISH) are ideal for this purpose as they allow subnuclear localisation via microscopy, and permit estimates of gene copy number by subnuclear spot intensity or number (Mark et al. Exp Mol Pathol 66:170-8 (1999)). Quantitative analysis of DNA for amplifications of Marker genes can also be carried out by Southern analysis, a method which is widely known to those skilled in the art (Sambrook et al., *supra*). DNA can be extracted from clinical material using established methods (Sambrook et al., *supra*). The methods of the invention can therefore also be directed to measuring genomic DNA levels of one or more of the identified Markers.

If a Marker mRNA encodes a secreted protein, that protein is likely to be present in body fluids. Proteins, especially secreted proteins, can be quantified in blood, serum, urine, semen and other fluids. Specific antibodies can often detect abundant proteins in ELISA tests on body fluid samples without enrichment. Prostate-specific antigen (PSA) falls into this category. PSA is an important Marker produced by prostatic epithelial cells and almost always expressed in prostate cancer, though not exclusively so. Clinically significant levels of proteins such as PSA are defined by appropriate studies, and protein levels are typically given as ng per mL of sample. Serum protein tests with greater specificity for cancer are needed. Detection of rare proteins may require that the protein is concentrated by e.g. precipitation. Thus, in a further embodiment of the invention diagnosis or prognosis or prostate disorder, or stage monitoring of the prostate disorder or therapeutic efficacy assessment is performed by testing for aberrant levels of one or more Marker proteins of the invention, which Marker protein is a secreted protein, in a bodily fluid.

The inventors predict that Markers 9, 19 and 24 (at least) are secreted proteins.

All the essential materials and reagents required for detecting or monitoring prostate disorder Markers in a test sample may be assembled together in a kit. Such a kit may comprise one or more diagnostic cDNA probes or oligonucleotide primers together with control probes/primers. The kit may contain probes immobilised on a microarray substrate such as a filter membrane or silicon-based substrate. The kit may also comprise samples of total RNA derived from tissues of various physiological states, such as normal, BPH, confined tumour and metastatic tumour, for example, to be used as controls. The kit may also comprise appropriate packaging and instructions for use in the methods of the invention.

Elmer 7700). Preferred oligonucleotide primers for detection of Marker mRNAs are selected from Markers 1-26.

In another embodiment the kit comprises one or more antibodies specific for one or other of the Markers identified herein for use in immunohistochemical analysis.

In another embodiment the kit is an ELISA kit comprising one or more antibodies specific for one or other of the Markers identified herein.

In another aspect of the invention, one or more of the 26 Markers can also be used in biochemical assays to identify agents which modulate the activity of the Marker proteins. The design and implementation of such assays will be evident to the biochemist of ordinary skill. The protein, particularly if it is a biochemical enzyme, may be used to turn over a convenient substrate whilst incorporating or losing a labelled component to define a test system. Test compounds are introduced into the test system and measurements made to determine their effect on enzyme activity. Such assays are useful to identify inhibitors of the enzyme which may prove valuable as therapeutic agents.

The inventors believe that Marker 19 (Factor V) is an example of a suitable biochemical enzyme that can be used in a suitable biochemical assay to identify modulators.

In a further aspect of the invention, each of the Markers can be used to characterise cell cultures in a screen for therapeutic agents, such as a high throughput screen. Effects of test compounds may be assayed by changes in mRNA or protein of any of Markers 1-26. As described above, cells (i.e. mammalian, bacterial etc) can be engineered to express one of the Markers identified herein.

Thus, according to a further aspect of the invention there is provided a method of testing potential therapeutic agents for the ability to suppress a prostate disorder phenotype comprising contacting a test compound with a cell engineered to express one of the Markers identified herein; and determining whether said test compound suppressed expression of the Marker.

Thus, according to a further aspect of the invention there is provided a screening assay or method for identifying potential anti-prostate disorder therapeutic compounds comprising contacting an assay system capable of detecting the effect of a test compound against expression levels of one or more of the Markers selected from the group consisting of: Marker 1 - Marker 26, with a test compound and assessing the change in expression level of the particular Marker under study.

Further features of the invention include:

A method of treatment of a patient suffering from a prostate disorder, comprising administration to the patient of a compound capable of reducing the transcription or expression of any one of Markers 1 - 26.

A method of treatment of a patient suffering from a prostate disorder, comprising administration to the patient an antisense nucleic acid molecule targeted against the mRNA of any one of Markers 1 - 26.

Use of an antisense nucleic acid molecule or an antibody directed against any one of Markers 1 - 26, in the manufacture of a medicament for treating a prostate disorder.

Each aspect of the invention involves detection or use of one or more of Markers 1 - 26. A preferred sub-group of Markers are those selected from the group consisting of Markers: 8, 13, 14, 15, 17, 18, 19, 20, 22, 24, 25 and 26.

The invention will be further described by way of the following non-limiting examples and figures in which data illustrating over-expression of markers is included:

Figure 1. Filter microarray data showing over-expression of Markers 1- 18 and 20-25 in prostate cancer relative to normal prostate. Overexpression value for Marker 26 (not shown) was found to be at the same level as Marker 13. Values given are mean expression level for 9 prostate cancer samples in those samples where over-expression of 1.5-fold or more was detected. Expression is given relative to normal prostate level (i.e. compared to the mean of 3 normal prostate datasets. These 3 datasets comprised 2 different RNA mixtures, each with at least 10 normal sample components).

Figure 2. Filter microarray data showing over-expression of Markers 1-5, 7-12, 17, 18, 20 and 23 in prostate cancer relative to BPH. Values given are mean expression level for 9 prostate cancer samples in those samples where over-expression of 1.5-fold or more was detected. Expression is given relative to BPH level (i.e. compared to the mean of 13 BPH datasets).

Example 1: Identification and Evaluation of Markers of Prostate Disease by cDNA Microarray Analysis

The microarray analysis protocol described in the following example was developed as a means to determine the relative abundances of mRNA species that are expressed in various tissues. Microarray analysis was used to identify differentially expressed RNA species

treated with DNAse I (also Gibco BRL), again as recommended by the manufacturer. For 1st strand label synthesis, 20µg total RNA was reverse transcribed into cDNA in the presence of radiolabelled nucleotide (^{33}P), using Superscript II enzyme (Gibco BRL) according to manufacturer's instructions. Labelled RNA was purified using GFX columns (Amersham Pharmacia Biotech, St. Albans, UK) according to manufacturers instructions, and added to 10 ml/filter of Church hybridisation buffer described below.

Prior to hybridisation, filters were wetted briefly in 2xSSC, then incubated in 5ml prewarmed Church hybridisation solution (0.5M sodium phosphate, 7% SDS, 1mM EDTA, pH 7.2) at 65°C for 2-6 hours. Probe was denatured at 100°C for 5 min, placed on ice for at least 5 min, then added to 5ml fresh Church hybridisation buffer in the presence of the filter, and mixed by swirling the bottle with the lid on for even probe distribution. Hybridisation was carried out at 65°C for 12-16 hours. Filters were rinsed briefly in pre-warmed Church wash solution (40mM sodium phosphate, 1% SDS, pH 7.2), then incubated twice in the same solution at 65°C for 20 min. Filters were drained briefly, wrapped in Saran wrap (Dow Chemical Company, USA) and exposed to phosphor screens for 3-6 days (screens and cassettes supplied by Molecular Dynamics, CA). The screens were then scanned using a Storm 830 phosphorimager (Molecular Dynamics). Array Vision software (Incyte Pharmaceuticals, CA, USA) was used to visualise the hybridisation images and generate quantitative numbers for each spot.

Typically, for comparison of data from different arrays, data is generated as a value relative to an internal standard for each array. Following export of ArrayVision data into Microsoft Excel format, spot measurements for data were normalised to housekeeping gene hybridisation signals known to be constant or relatively constant, using a simple Excel macro. A total of 38 spot values (from 16 different clones) representing 12 different housekeeping genes were used to generate an average housekeeping measurement. The gene and EMBL accession number of the housekeeping genes used are listed in Table 1. Background values were not subtracted, but to eliminate spurious low-level signals indistinguishable from noise, a local background value from 3 blank spots out of every 7x7 array was taken, and every value <2-fold this value was ignored.

To generate representative values for normal prostate, data from 3 array hybridisation experiments to normal prostate RNA was normalised to the mean housekeeping gene value, then the mean of these 3 values was taken as the control dataset.

histopathological analysis. **Table 2** lists the 26 genes we have identified as being Markers for prostatic disease.

Control number	Gene	NCBI genInfo identifier	Spots on Array
1	Human ribosomal protein S14 gene	g337498	2
2	Human ribosomal protein S14 gene	g337498	2
3	Human acidic ribosomal phosphoprotein P0	g190231	2
4	Human alpha-tubulin mRNA	g340020	2
5	Human mRNA for TEF-5 protein	g1848081	2
6	Human transcriptional enhancer factor	g339440	2
7	Human transcription factor RTEF-1	g1561727	2
8	Human hypoxanthine phosphoribosyltransferase	g184349	2
9	Human hypoxanthine phosphoribosyltransferase	g184349	2
10	Human mRNA for transcription factor, TEF3	g1403337	2
11	Human acidic ribosomal phosphoprotein P1	g190233	2
12	Human mRNA for ribosomal protein L19	g36127	4
13	Human hH3.3B gene for histone H3.3	g761715	4
14	Human pancreatic phospholipase A-2 (PLA-2)	g190008	2
15	Human mRNA for ribosomal protein L19	g36127	4
16	Human acidic ribosomal phosphoprotein P1	g190233	2

Table 1. Names and identities of control housekeeping genes used to normalise microarray data. This set was identified from a wider set, by comparison of variability of candidate housekeeping gene signals over 13 filter hybridisation experiments as a group showing minimal variability.

prostate cancer compared to BPH, or in the case of Marker 19, not detected at all in normal prostate. In addition, Markers 6, 13, 14 and 25 were not detected in BPH.

Example 2: Quantitative RT-PCR test

RT-PCR is a convenient method for assaying the relative abundance of mRNA species expressed in various cells, tissues and organs.

Reproducibly accurate amplification of an mRNA can be achieved by firstly synthesising a DNA template from the mRNA using a reverse transcriptase, then by PCR using a heat-stable DNA polymerase (e.g. Taq polymerase). An optimised amplification protocol is essential for quantitation of nucleic acids by PCR because small differences in efficiency of the reaction can greatly influence the reaction rates, with a subsequent effect on PCR yield. Optimisation of PCR is not usually problematic (refer to: Koehler *supra,* and references therein). Quantitation of RT-PCR products can be done while the reaction products are building up exponentially, preferably following each round of amplification (known as real-time PCR). For quantitation, analysis is carried out by reference to one or more housekeeping genes which are also amplified by RT-PCR, often in a different reaction tube. Quantitation of RT-PCR product may be undertaken, for example, by gel electrophoresis visual inspection or image analysis, HPLC (Koehler et al., *supra*) or by use of suitable detection methods such as described above.

In practise, specific pairs of oligonucleotide primers corresponding to Markers 1-26 can be used to trigger amplification of specific marker mRNAs. The Marker mRNA is quantitated relative to a standard housekeeping mRNA using a real-time RT-PCR assay. For the purpose of this example, the housekeeping gene is ribosomal protein L19, but any such gene could be used.

Firstly, RNA is isolated from a clinical or tissue sample as in example 1. Aliquots of RNA are used in a reverse transcription reaction using random hexamers or oligo-dT as primers, and for example Moloney Murine Leukaemia Virus (MuLV) or Avian Myeloblastosis Virus (AMV) reverse transcriptase. A typical 20ml reverse transcription reaction comprises the following: 10 mM each dNTP, 50 mM Tris-HCl, pH8.3, 50 mM KCl, 50 mM dithiothrietol, 0.5 mM spermidine, 10ng oligo-dT primer, up to 1µg RNA, dissolved in DEPC-treated water. This mixture is incubated at 65°C for 15 min, then chilled on ice for 5 min. 5 units of AMV reverse transcriptase and 10 units of RNAse inhibitor are added. The

Example 3: Generation of antibody specific to Protein encoded by Markers 1-26

Antibody specific to Markers 1-26 could be used to detect protein products of these Markers in tissue samples and body fluids. Ideally for diagnosis, easily accessible samples are required, involving minimally invasive techniques. Typically, these samples include blood, plasma, serum, saliva, urine and semen. These fluids are far removed from prostate cells and, in a normal individual, would not be expected to contain mRNA of prostatic origin. However, protein may be transported across cell membrane barriers into body fluids. Prostate-specific antigen is one such protein, and is easily assayed by a simple blood test using specific antibody against the protein. For new markers, e.g. Markers 1-26, it may be necessary to generate novel antibody. This example describes a method of generating protein in vitro, which can then be used to raise antibody.

One way to generate protein is to express the cloned gene in *E. coli*. This requires that the cDNA, e.g. Markers 1-26 is cloned into a vector capable of expressing the encoded protein in the bacterial host. Optionally, it may be desirable to incorporate a molecular tag into the protein so that it can be easily purified. One such tag in wide use is the 6xHis tag. A protein with this tag is easily isolated from a cell culture extract by affinity chromatography, and can be used in relatively pure form to inoculate an animal for antibody generation.
Having subcloned the cDNA or cDNA fragment into an appropriate vector, e.g. pHAT10, 11, 12 or 20 (Clontech, Palo Alto, CA) and confirmed the integrity of the insert such as sequence and orientation. The subclone is cultured on a suitable scale, cells are harvested by centrifugation and then lysed, and the protein extract incubated with TALON resin according to manufacturer's instructions (Clontech, Palo Alto, CA). The resin is washed and recombinant protein is eluted by adjusting the pH or imidazole concentration. If desired, the 6xHIS tag can be enzymatically removed using a specific protease (Clontech). To generate a specific antibody, the purified protein is injected into a host animal, usually rabbit, sheep or goat. After boost injections, the serum is periodically collected and tested for antibody. Polyclonal antibody can be purified form the serum using standard techniques, or the animal's spleen can be harvested for the production of hybridomas. Techniques for the production of antibodies, both polyclonal and monoclonal, are well known to those skilled in the art (Catty D (ed.) Antibodies: A Practical Approach. Vol 1 (1988), Vol2 (1989)).

A hybridisation probe can be generated from a synthetic oligonucleotide or a dephosphorylated restriction fragment sequence by addition of a radioactive 5' phosphate group from [γ-^{32}P]ATP by the action of T4 polynucleotide kinase. 20 pmoles of the oligonucleotide are added to a 20µl reaction containing 100mM Tris, pH7.5 , 10mM $MgCl_2$, 0.1mM spermidine, 20mM dithiothreitol (DTT), 7.55µM ATP, 55µCi [γ-^{32}P]ATP and 2.5u T4 polynucleotide kinase (Pharmacia Biotechnology Ltd, Uppsala, Sweden). The reaction is incubated for 30 minutes at 37°C and then for 10 minutes at 70°C prior to use in hybridisation. Methods for the generation of hybridisation probes from oligonucleotides or from DNA and RNA fragments are described in Chapters 11 and 10 respectively in Sambrook et al. (*ibid*). A number of proprietary kits are also available for these procedures.

Filter preparation

The sample DNA could be isolated and run on an agarose gel and Southern blotted onto a nitrocellulose or nylon filter using standard techniques.

Hybridisation conditions

Filters containing the nucleic acid are pre-hybridised in 100ml of a solution containing 6x SSC, 0.1%SDS and 0.25% dried skimmed milk (Marvel™) at 65°C for a minimum of 1 hour in a suitable enclosed vessel. A proprietary hybridisation apparatus such as model HB-1 (Techne Ltd) provides reproducible conditions for the experiment.

The pre-hybridisation solution is then replaced by 10ml of a probe solution containing 6xSSC, 0.1% SDS, 0.25% dried skimmed milk (e.g. Marvel™) and the oligonucleotide probe generated above. The filters are incubated in this solution for 5 minutes at 65°C before allowing the temperature to fall gradually to below 30°C. The probe solution is then discarded and the filters washed in 100ml 6xSSC, 0.1% SDS at room temperature for 5 minutes. Further washes (1-3) are then made in fresh batches of the same solution at 30°C and then, optionally, in 10°C increments up to 60°C for 5 minutes per wash.

After washing, the filters are dried and used to expose an X-ray film such as Hyperfilm™ MP (Amersham International) at -70°C in a light-tight film cassette using a fast tungstate intensifying screen to enhance the photographic image. The film is exposed for a suitable period (normally overnight) before developing to reveal the photographic image of the radio-active areas on the filters. Related nucleic acid sequences are identified by the presence of a photographic image compared to totally unrelated sequences which should not

then chilled quickly on ice-water, is placed on the tissue section and a coverslip is sealed on to the slide using rubber solution. A suitable probe is a digoxygenin-labelled 300-600 base pair cDNA clone or fragment thereof, corresponding to one of Markers 1-26, prepared by one of several methods known to the art. The slides are then incubated in a sealed chamber for 16h at 37°C. The rubber solution is peeled away and coverslips are removed by immersion of the slides in 2x SSC, 0.1% SDS. Slides are washed 4 times for 5 min each at room temperature, and twice for 10 min at 65°C, in 2x SSC, 0.1% SDS, and rinsed briefly in 2x SSC. Hybridisation is detected as follows: Nonspecific binding is blocked by incubating the slides in blocking buffer (0.1M Tris-HCL pH 7.5, 0.1M NaCl, 2mM $MgCl_2$, 3% bovine serum albumin) for 10 min. Slides are then flooded with anti-digoxygenin-conjugated alkaline phosphatase (1/500 dilution of stock; Boehringer, Mannheim, Germany) in blocking buffer, and incubated 2 h at room temperature, then washed in blocking buffer 3 times for 3 min each. Slides are placed in buffer 2 (0.1 M Tris-HCl pH 9.5, 0.1 M NaCl, 50 mM $MgCl_2$) for 10 min, then immersed in substrate buffer (100mL buffer 2 containing 25 mg levamisole, plus 35 mg nitroblue tetrazolium chloride dissolved in 277 mL 70% dimethylformamide, plus 17 mg 5-bromo-4-indoyl-phosphate dissolved in 222 mL 100% dimethylformamide) for 10-30 min in the dark. The slides are then immersed in 20 mM Tris-HCl pH7.5, 5 mM EDTA for 5 min, rinsed with tap water for 5-10 min, mounted in aqueous mountant and examined under a microscope.

Example 7. Screening for therapeutic agents capable of inhibiting expression of any of Markers 1-26.

Markers 1-26 may be employed in a process for screening compounds which either inhibit, promote or modulate the expression of Markers 1-26. Examples of potential Marker 1-26 agonists are small molecules such as organic molecules or peptides, antibodies or oligonucleotides which bind to Markers 1-26 and inhibit expression or activity. One assay for therapeutic agents uses cultured cells, and measures transcript abundance using a microarray as described in Example 1. Typical prostate cell lines such as LNCaP, PC-3 and DU145 are available from the European Collection of Cell Cultures, Salisbury UK. Cell culture is a standard technique well known to those practised in the art. Briefly, in one suitable example, a seed vial of LNCaP cells is inoculated into 500mL culture flasks containing 20-50ml RPMI medium containing 10% fetal calf serum and 1% glutamate (all components Gibco, UK).

Claims:

1. A method for diagnosing or prognosing or monitoring a prostate disorder comprising testing a biological sample from an individual for aberrant levels of one or more of the Markers selected from the group consisting of: Marker 1 to Marker 26.

2. A method as claimed in claim 1, wherein the Marker(s) are selected form the group consisting of Markers: 8, 13, 14, 15, 17, 18, 19, 20, 22, 24, 25 and 26.

3. A method for distinguishing prostate cancer from BPH comprising testing a biological sample for aberrant levels of one or more of the Markers selected from the group consisting of Marker 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, 18, 20, 23, 25 and 26.

4. A method as claimed in any of claims 1-3 wherein mRNA transcript levels and/or protein levels of the Marker(s) is/are measured.

5. A method as claimed in claim 4, wherein mRNA transcript levels are measured using reverse-transcriptase polymerase chain reaction (RT-PCR).

6. A method for measuring Marker mRNA transcript levels as claimed in claim 4, which method involves the use of one or more oligonucleotide probes each capable of selectively hybridising to nucleic acid of a Marker of interest to determine the expression level of said Marker of interest.

7. A method for measuring Marker protein levels as claimed in claim 4, which method involves the use of one or more antibodies each capable of selectively binding to a Marker protein or protein fragment of interest to determine the expression level of said Marker of interest.

8. Use of an antibody selective for a Marker protein selected from the group consisting of: Marker 1 - Marker 26, in an assay to monitor therapeutic efficacy.

9. A diagnostic kit for diagnosing or prognosing or monitoring a prostate disorder comprising, one or more diagnostic probe(s) and/or one or more diagnostic primer(s) and/or one or more antibodies capable of selectively hybridising or binding to one or more of the Markers 1 -26.

10. A screening assay for identifying potential anti-prostate disorder therapeutic compounds comprising contacting an assay system capable of detecting the effect of a test compound on expression levels of one or more of the Markers selected from the group consisting of: Marker 1 - Marker 26, with a test compound and assessing the change in expression level of the particular Marker under study.

11. A method of testing potential therapeutic agents for the ability to suppress a prostate disorder phenotype comprising contacting a test compound with a cell engineered to express one of the Markers identified herein; and determining whether said test compound suppressed expression of the Marker.

12. A compound or agent identified by the screening assay according to claim 10 or the method according to claim 11.

13. A method of treatment of a patient suffering from a prostate disorder, comprising administration to the patient of a compound capable of reducing the transcription or expression of any one of Markers 1 - 26.

Figure 1

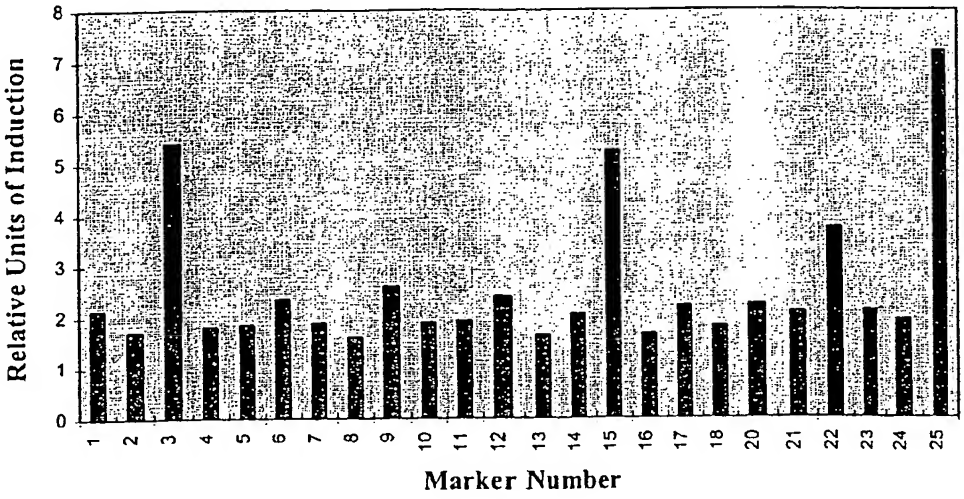
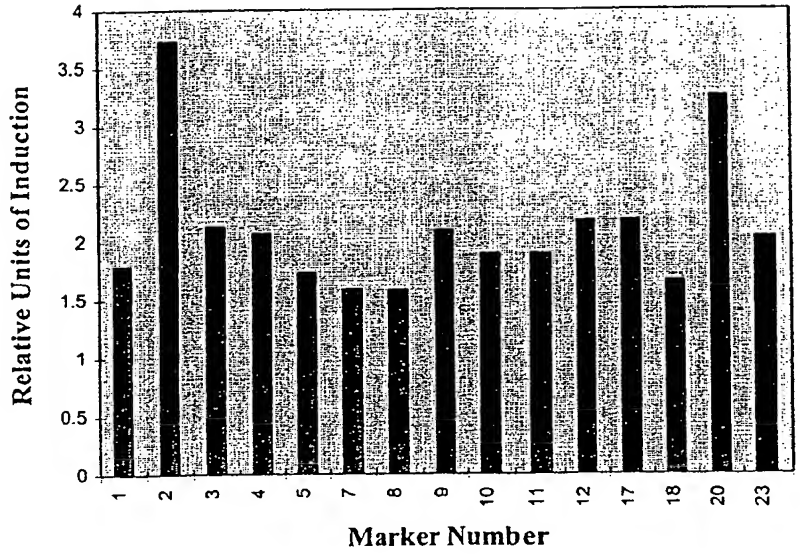


Figure 2



cctccaccct ggtccgcctc cctggcctgg ctcaaccgag tgcctccgac ccccctcctc 840
agccctcccc cacccacagg cccagcctcc tcggtctcct gtctcgttgc tgcttctgcc 900
tgtgctgtgg gggagagagg ccgcagccag gcctctgctg ccctttctgt gccccccagg 960
ttctatctcc ccgtcacacc cgaggcctgg cttcaggagg gagcggagca gccattctcc 1020
aggccccgtg gttgcccctg gacgtgtgcg tctgctgctc cggggtggag ctggggtgtg 1080
ggatgcacgg cctcgtgggg gccgggccgt cctccagccc cgctgctccc tggccagccc 1140
ccttgtcgct gtcggtcccg tctaaccatg atgccttaac atgtggagtg taccgtgggg 1200
cctcactagc ctctactccc tgtgtctgca tgagcatgtg gcctccccgt cccttccccg 1260
gtggcgaacc cagtgaccca gggacacgtg gggtgtgctg gtgctgctcc ccagcccacc 1320
aatgcctggc cagcctgccc ccttccctgg acagggctgt ggagatggct ccggcggctt 1380
ggggaaagcg aaattgccaa cactcaagtc acctcagtac catccaggag gctgggtatt 1440
gtcctgcctc tgccttttct gtctcagcgg cagtgcccag agcccacacc cccccaagag 1500
ccctcgatgg acaggcctga cccaccccac ctggggccag ccaggagccc cgcctgggcc 1560
atcagtattt attgcctccg tccgtgccgt ccctgggcca ctggctggcg cctcttcccc 1620
cagcctctca gtgccaccac ccccggcagc cttccctgac ccagccagga caaacaaggg 1680
accaagtgca cacattgctg agagccgtct cctataggtc ccccgcccca tccccggtgt 1740
tggtgttgtg tctgccaggc tcaggcagag gcgcctgtcc ctgcttcttt tctgaccggg 1800
aaataaatgc ccctgaagg 1819

<210> 3
<211> 996
<212> DNA
<213> Homo sapiens

<400> 3
agcttctctg atccctagta taaacacttc agtgttcccc tttcagtctt actactttga 60
ccgcgatgat gtggctttga agaactttgc caaatacttt cttcaccaat ctcatgagga 120
gagggaacat gctgagaaac tgatgaagct gcagaaccaa cgaggtggcc gaatcttcct 180
tcaggatatc aaggtgaaca aaagatccta ggggtgtcat acttcatcat ctggcagtgt 240
tcgggtatca gaaatcactt aaactagcaa ttgcccttat aaagtgatga tacactgggc 300
ttttgccttt tgtgcttttt taggcttacc atctaaacta aattaggcaa atagtaatgt 360
ccctttgcc aaaacgtggt ggttagagat gatgggcttg ctgacttcta ggttagttgg 420
tagagatgca ttaacctatt ctcattcaga aaccagactg tgatgactgg gagagcgggc 480
tgaatgcaat ggagtgtgca ttacatttgg aaaaaaatgt gaatcagtca ctactggaac 540
tgcacaaact ggccactgac aaaaatgacc cccatgtgag tattggaacc ccaggaaata 600
aatggaggaa atcatttgcc ttagggattg ggaaagctgc ccactaactg tcttgcccca 660
ttgttttgca gttgtgtgac ttcattgaga cacattacct gaatgagcag gtgaaagcca 720
tcaaagaatt gggtgaccac gtgaccaact tgcgcaagat gggagcgccc gaatctggct 780
tggcggaata tctctttgac aagcacaccc tgggagacag tgataatgaa agctaagcct 840
cgggctaatt tccccatagc cgtggggtga cttccctggt caccaaggca gtgcatgcat 900
gttggggttt cctttacctt ttctataagt tgtaccaaaa catccactta agttctttga 960
tttgtaccat tccttcaaat aaagaaattt ggtacc 996

<210> 4
<211> 2353
<212> DNA
<213> Homo sapiens

<400> 4
gggttgagca caaggaggag aacgaccaca aagtcttcta cgggggtgac ctgaaagtgg 60
actgtgtggc caccgggctt cccaatcccg agatctcctg gagcctccca gacgggagtc 120
tggtgaactc cttcatgcag tcggatgaca gcggtggacg caccaagcgc tatgtcgtct 180
tcaacaatgg gacactctac tttaacgaag tggggatgag ggaggaagga gactacacct 240
gctttgctga aaatcaggtc gggaaggacg agatgagagt cagagtcaag gtggtgacag 300
cgcccgccac catccggaac aagacttgct tggcggttca ggtgccctat ggagacgtgg 360
tcactgtagc ctgtgaggcc aaaggagaac ccatgcccaa ggtgacttgg ttgtccccaa 420
ccaacaaggt gatccccacc tcctctgaga agtatcagat ataccaagat ggcactctcc 480
ttattcagaa agcccagcgt tctgacagcg gcaactacac ctgcttggtc aggaacagcg 540
cgggagagga taggaagacg gtgtggattc acgtcaacgt ccagccaccc aagatcaacg 600
gtaaccccaa ccccatcacc actgtgcggg agatagcagc cgggggcagt cggaaactga 660
ttgactgcaa agctgaaggc atccccaccc cgagggtgtt atgggctttt cccgagggtg 720
tggttctgcc agctccatac tatggaaacc ggatcactgt ccatggcaac ggttccctgg 780

<211> 5086
<212> DNA
<213> Homo sapiens

<400> 6
gcctgcgccg cgcagcccgc ctcgggcggg aggcgggagg cgggaggccg ggcccaggcc 60
ggggcagccc ctccaccgca ccgtcctggg ccggtgccca ggtccgagtc gccttccgcc 120
tgccccccgc gccaatcccc cgccgcggca gccccagcca ggtcccgccg ccaggccggc 180
tcccgccccc gctccgcccc cggagccgca gccccgcccg ccatcgccgt cgccatgttg 240
tggctcccgc agccggcgct ggggacgcgc gcggccgaga ctctggcctg cagtcgccgc 300
cgccgccgcc aggtggtgtt tggactctag accatgtgcc taggtagaag tttttccttt 360
ctccgcagct ctgctccgct agcaacgctc gccacaccct tgttttgaga tcctctctaa 420
ggagcggaga gtttaatagg caagaaggaa gggagaagac agaaggaaga cgctcccccg 480
tacggagaca gagggagggg gggctccaaa gccgaaagag gaggtcccta cctgccacgg 540
ataccagtca gcccttgcca gcatccagcc atgggggata tgaaaacccc agattttgat 600
gaccttctgg ctgcctttga catcccagac cccaccagcc ttgatgccaa ggaggccatc 660
cagacaccca gtgaggagaa tgagagtccc ctcaaacctc caggcatatg tatggatgaa 720
agtgtgtcct tgtctcactc aggatcagcc cccgatgtgc cggccgtgag tgtcattgtc 780
aagaacacca gccgccagga gtcatttgaa gcggagaaag accacattac tcccagtctc 840
ctacacaatg gattccgggg ctcagatctg cctccagatc cccacaactg tgggaaattt 900
gattctactt ttatgaatgg agacagtgcc aggagtttcc ctggcaaact ggagcctccc 960
aagtcagagc cattacccac cttcaaccag ttcagtccaa tctccagccc agaacctgag 1020
gatcccatca aagataacgg atttgggata aagcccaaac actctgacag ttatttccca 1080
ccccctcttg ggtgcggggc tgtgggaggc ccagtcctgg aggctctggc taagtttccg 1140
gttccagagc tgcatatgtt tgatcatttt tgtaagaaag aacccaagcc agaacccctg 1200
cccttgggga gccagcagga acacgagcaa agtgggcaga acacagtgga acctcacaag 1260
gatccggatg ccactcgatt cttcggggaa gctttggagt tcaacagcca tcctagcaac 1320
agtattggag agtccaaggg gcttgcccgg gagcttggta cctgctcatc agtccccctt 1380
aggcagcgtc taaagccagc tcattccaag ctgtcctctt gtgtggcagc cttggtggcc 1440
ttgcaggcca aaagagtggc tagtgtcact aaggaggatc agcctggcca cacaaaggat 1500
ctctcagggc ccactaaaga gagttctaaa ggtagcccca aaatgcccaa gtcaccaaag 1560
agtccccgga gcccctctgga ggccactaga aaaagtatca agccatcgga cagccctcgt 1620
agcatctgca gtgacagcag cagcaaaggc tcaccgtctg tggctgccag ctccccacca 1680
gcaattccca aagtgagaat caaaaccatt aagacatcat caggggaaat caaacggact 1740
gtcacaagga tcctgccaga tcctgatgat ccaagtaagt cccctgttgg gtcacctcta 1800
gggagcgcca ttgcagaggc ccccagcgag atgccagggg atgaggtgcc tgtggaagag 1860
cactttcctg aggcaggcac aaattcaggg agcccccagg gggccaggaa aggggacgag 1920
agcatgacaa aggccagtga ctcgtcatct cccagctgca gttctgggcc ccgggtccca 1980
aaggggctg ccccaggctc acagacaggc aagaagcaac agagcacagc actgcaggca 2040
tccaccctgg cccctgccaa cctcctgccc aaagccgtgc acttggccaa cctgaacctc 2100
gtcccccaca gtgttgctgc atcagtgaca gccaagtctt cagtgcaaag acggagccag 2160
ccacagctta cacaaatgtc ggtgcccctg gtccaccagg tgaaaaaggc tgccccactg 2220
attgtagagg tcttcaacaa ggtccttcac agctccaacc ccgtgccctt ctatgcgcca 2280
aatctcagcc cgcctgcgga cagcaggatc cacgtgccgg ccagtgggta ctgctgcctg 2340
gagtgtggag acgcatttgc cttagagaag agcctgagcc agcactatgg ccggcggagc 2400
gtccacattg aggtactgtg cacactgtgc tccaagacgc tgctcttctt caacaagtgc 2460
agcctgctcc ggcacgcccg tgaccacaag agcaaggggc tcgtcatgca gtgttcccag 2520
ctgctggtga agcctatctc tgcggaccaa atgttcgtgt cggcccctgt gaactccacg 2580
gcaccagcag ccccagcccc ttcatcctct cccaaacatg gcctcacttc gggcagtgcc 2640
agtccccctc ctccagcctt gccactctac ccagaccctg tgaggctcat ccggtactca 2700
atcaagtgtc ttgaatgtca caagcagatg cgggactaca tggtcctggc tgcacatttc 2760
cagaggacaa cagaggagac agaggggctg acctgccagg tatgccagat gctgctgccc 2820
aaccagtgca gtttctgtgc ccaccagcgg attcatgcac acaagtcccc ctactgctgc 2880
ccggagtgtg gggtcctctg ccgctctgcc tacttccaga cccatgtaaa ggagaattgc 2940
ctgcactatg cccgcaaggt gggctacagg tgcatccact gtggtgtcgt ccacctgacc 3000
ttggccttgc tgaaaagcca catccaggag cgacactgcc aggttttcca caaatgtgca 3060
ttctgcccca tggccttcaa gactgccagc agcactgcag accacagtgc cacccagcac 3120
cccacccagc cccacagacc ctcccagctc atttataagt gctcctgtga aatggtcttc 3180
aacaagaaga ggcacattca gcagcatttt taccagaatg tcagcaagac gcaggtgggc 3240
gtcttcaagt gccctgagtg cccactcttg ttcgtgcaga agccggagtt gatgcaacac 3300
gtcaagagca cccacggtgt tccccgaaat gtggacgagc tgtcaaacct ccagtcttca 3360
gcggacacat cctcaagccg ccctggctct cgagttccca ctgagccacc agccactagt 3420
gtggctgctc ggagcagctc cctgccttct ggccgctggg gtaggcctga agcccaccgc 3480
agggtggaag ccaggccgcg gctgaggaac actggctgga cctgccagga gtgccaggag 3540

<400> 8
ttttttttt ttttttgtcc gacatcttga gacgaggctg cggtgtctgc tgctattctc 60
cgagcttcgc aatgccgcct aaggacgaca agaagaagaa ggacgctgga aagtcggcca 120
agaaagacaa agacccagtg aacaaatccg ggggcaaggc caaaaagaag aagtggtcca 180
aaggcaaagt tcgggacaag ctcaataact tagtcttgtt tgacaaagct acctatgata 240
aactctgtaa ggaagttccc aactataaac ttataacccc agctgtggtc tctgagagac 300
tgaagattcg aggctccctg gccagggcag cccttcagga gctccttagt aaaggactta 360
tcaaactggt ttcaaagcac agagctcaag taatttacac cagaaatacc aagggtggag 420
atgctccagc tgctggtgaa gatgcatgaa taggtccaac cagctgtaca tttggaaaaa 480
taaaacttta ttaaatc 497

<210> 9
<211> 2056
<212> DNA
<213> Homo sapiens

<400> 9
ggaaccgcgg ctgctggaca agaggggtgc ggtggatact gacctttgct ccggcctcgt 60
cgtgaagaca cagcgcatct ccccgctgta ggcttctccc acagaacccg tttcgggcct 120
cagagcgtct ggtgagatgc tgttgccgct gctgctgctg ctacccatgt gctgggccgt 180
ggaggtcaag aggccccggg gcgtctccct caccaatcat cacttctacg atgagtccaa 240
gcctttcacc tgcctggacg gttcggccac catcccattt gatcaggtca acgatgacta 300
ttgcgactgc aaagatggct ctgacgagcc aggcacggct gcctgtccta atggcagctt 360
ccactgcacc aacactggct ataagccccc gtatatcccc tccaaccggg tcaacgatgg 420
tgtttgtgac tgctgcgatg gaacagacga gtacaacagc ggcgtcatct gtgagaacac 480
ctgcaaagag aagggccgta aggagagaga gtccctgcag cagatggccg aggtcacccg 540
cgaagggttc cgtctgaaga agatccttat tgaggactgg aagaaggcac gggaggagaa 600
gcagaaaaag ctcattgagc tacaggctgg gaagaagtct ctggaagacc aggtggagat 660
gctgcggaca gtgaaggagg aagctgagaa gccagagaga gaggccaaag agcagcacca 720
gaagctgtgg gaagagcagc tggctgctgc caaggcccaa caggagcagg agctggcggc 780
tgatgccttc aaggagctgg atgatgacat ggacgggacg gtctcggtga ctgagctgca 840
gactcacccg gagctggaca cagatgggga tggggcgttg tcagaagcgg aagctcaggc 900
cctcctcagt ggggacacac agacagacgc cacctctttc tacgaccgcg tctgggccgc 960
catcagggac aagtaccggt ccgaggcact gcccaccgac cttccagcac cttctgcccc 1020
tgacttgacg gagcccaagg aggagcagcc gccagtgccc tcgtcgccca cagaggagga 1080
ggaggaggag gaggaggagg aagaagaggc tgaagaagag gaggaggagg aggattccga 1140
ggaggcccca ccgccactgt caccccccgca gccggccagc cctgctgagg aagacaaaat 1200
gccgccctac gacgagcaga cgcaggcctt catcgatgct gcccaggagg cccgcaacaa 1260
gttcgaggag gccgagcggt cgctgaagga catggaggag tccatcagga acctggagca 1320
agagatttct tttgactttg gccccaacgg ggagtttgct tacctgtaca gccagtgcta 1380
cgagctcacc accaacgaat acgtctaccg cctctgcccc ttcaagcttg tctcgcagaa 1440
acccaaactc gggggctctc ccaccagcct tggcacctgg ggctcatgga ttggccccga 1500
ccacgacaag ttcagtgcca tgaagtatga gcaaggcacg ggctgctggc agggccccaa 1560
ccgctccacc accgtgcgcc tcctgtgcgg gaaagagacc atggtgacca gcaccacaga 1620
gcccagtcgc tgcgagtacc tcatggagct gatgacgcca gccgcctgcc cggagccacc 1680
gcctgaagca cccaccgaag acgaccatga cgagctctag ctggatgggc gcagagaacc 1740
tcaagaaggc atgaagccag cccctgcagt gccgtccacc cgcccctctg ggcctgcctg 1800
tggctctgtt gccctcctct gtggcggcag gacctttgtg gggcttcgtg ccctgctctg 1860
gggcccaggc ggggctggtc cacattccca ggccccaaca gcctccaaag atgggtaaag 1920
gagcttgccc tccctgggcc ccccaccttg gtgactcgcc ccaccacccc cagccctgtc 1980
cctgccaccc ctcctagtgg ggactagtga atgacttgac ctgtgacctc aatacaataa 2040
atgtgatccc ccaccc 2056

<210> 10
<211> 4577
<212> DNA
<213> Homo sapiens

<400> 10
gctacaatcc atctggtctc ctccagctcc ttctttctgc aacatgggga agaacaaact 60
ccttcatcca agtctggttc ttctcctctt ggtcctcctg cccacagacg cctcagtctc 120
tggaaaaccg cagtatatgg ttctggtccc ctccctgctc cacactgaga ccactgagaa 180

ctccttgaaa tacaatattc tcccagaaaa ggaagagttc ccctttgctt taggagtgca 4080
gactctgcct caaacttgtg atgaacccaa agcccacacc agcttccaaa tctccctaag 4140
tgtcagttac acagggagcc gctctgcctc caacatggcg atcgttgatg tgaagatggt 4200
ctctggcttc attccccctga agccaacagt gaaaatgctt gaaagatcta accatgtgag 4260
ccggacagaa gtcagcagca accatgtctt gatttacctt gataaggtgt caaatcagac 4320
actgagcttg ttcttcacgg ttctgcaaga tgtcccagta agagatctca aaccagccat 4380
agtgaaagtc tatgattact acgagacgga tgagtttgca atcgctgagt acaatgctcc 4440
ttgcagcaaa gatcttggaa atgcttgaag accacaaggc tgaaaagtgc tttgctggag 4500
tcctgttctc tgagctccac agaagacacg tgtttttgta tctttaaaga cttgatgaat 4560
aaacactttt tctggtc 4577

<210> 11
<211> 1651
<212> DNA
<213> Homo sapiens

<400> 11
ggagcccccg ccctgggatt cccaggtgtt ttcatttggt gatcagcact gaacacagaa 60
gagtcatgac ggagtttggg ctgagctggg ttttccttgt tgctattttt aaaggtgtcc 120
agtgtgaggt gcagctggtg gagtctgggg gagacttggt ccagcctggg gggtccctga 180
gactctcctg tgcagcctct ggattcacct tcagtagtta tgctatgcac tgggtccgcc 240
aggctccagg gaagggactg aaatatgttt caggtattag tagtaatggg cgtagaacat 300
attatgcaaa ttctgtgaag ggcagattca ccatctccag agacaattcc aagaacacgt 360
tgtatcttca aatgggcagc ctgagagctg aggacatggc tgtgtattac tgtgtgtccg 420
ggggaatcta tgatagtagt ggtccctttg actactgggg ccagggaacc ctggtcaccg 480
tctcctcagc atccccgacc agccccaagg tcttcccgct gagcctctgc agcacccagc 540
cagatgggaa cgtggtcatc gcctgcctgg tccagggctt cttcccccag gagccactca 600
gtgtgacctg gagcgaaagc ggacagggcg tgaccgccag aaacttccca cccagccagg 660
atgcctccgg ggacctgtac accacgagca gccagctgac cctgccggcc acacagtgcc 720
tagccggcaa gtccgtgaca tgccacgtga agcactacac gaatcccagc caggatgtga 780
ctgtgccctg cccagttccc tcaactccac ctaccccatc tccctcaact ccacctaccc 840
catctccctc atgctgccac ccccgactgt cactgcaccg accggccctc gaggacctgc 900
tcttaggttc agaagcgaac ctcacgtgca cactgaccgg cctgagagat gcctcaggtg 960
tcaccttcac ctggacgccc tcaagtggga agagcgctgt tcaaggacca cctgagcgtg 1020
acctctgtgg ctgctacagc gtgtccagtg tcctgccggg ctgtgccgag ccatggaacc 1080
atgggaagac cttcacttgc actgctgcct accccgagtc caagaccccg ctaaccgcca 1140
ccctctcaaa atccggaaac acattccggc ccgaggtcca cctgctgccg ccgccgtcgg 1200
aggagctggc cctgaacgag ctggtgacgc tgacgtgcct ggcacgcggc ttcagcccca 1260
aggacgtgct ggttcgctgg ctgcaggggt cacaggagct gccccgcgag aagtacctga 1320
cttgggcatc ccggcaggag cccagccagg gcaccaccac cttcgctgtg accagcatac 1380
tgcgcgtggc agccgaggac tggaagaagg gggacacctt ctcctgcatg gtgggccacg 1440
aggccctgcc gctggccttc acacagaaga ccatcgaccg cttggcgggt aaacccaccc 1500
atgtcaatgt gtctgttgtc atggcggagg tggacggcac ctgctactga gccgcccgcc 1560
tgtccccacc cctgaataaa ctccatgctc ccccaaaaaa aaaaaaaaaa aaaaaaaaaa 1620
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa a 1651

<210> 12
<211> 1120
<212> DNA
<213> Homo sapiens

<400> 12
acttcttcgc accagggaag ccccacccac cagaacgcca agatgtccag caagcgggcc 60
aaagccaagg ccaccaagaa gcggccacag cgggccacat ccaatgtctt cgcaatgttt 120
gaccagtccc agatccagga gtttaaggag gctttcaaca tgattgacca gaaccgtgat 180
ggcttcattg acaaggagga cctgcacgac atgctggcct cgctggggaa gaaccccaca 240
gacgaatacc tggagggcat gatgagcgag gccccggggc catacaactt caccatgttc 300
ctcaccatgt ttggggagaa gctgaacggc acggaccccg aggatgtgat tcgcaacgcc 360
tttgcctgct tcgacgagga atcctcaggt ttcatccatg aggaccacct ccggaagctg 420
ctcaccacca tgggtgaccg cttcacagat gaggaagtgg acgagatgta ccgggaggca 480
cccgttgata agaaaggcaa cttcaactac gtggagttca cccgcatcct caaacatggc 540
gccaaggata aacacgacta ggccatcccc agccccctga cacccagccc ccgccagtca 600

<210> 14
<211> 3612
<212> DNA
<213> Homo sapiens

<400> 14
aagcttaaag tgaagacttg ctgaaattat ttcagcattt tactcctggc cattttccat 60
aggaaggctc ctaatagtgt ctaactcctg aaagcagaaa gtgaccttca aaactgaccc 120
atcttccctg gaatctctct aagagctgag ggattagggt tcttagagat ttgtgtatct 180
ttgatggcac taaaatgaca tagagtttac atgtatacac ctggaaaagg tattaaatgt 240
tatcagatgg ccctcaaagg tctagaacat ggtagctgtg cattgaggaa acattctgca 300
aaaaccattg tctcagaaag aaacatccag ctctatatta attgtatgct cagacccatc 360
cctgctggtt aggtcagcat ttgtacttgg tttgcatatt ataccattgt gtgacatcaa 420
caatctgcaa gtcacatttg ccgggctgcc tgtgcagaat ggcagctgct gggcctagag 480
tctagtgcag gtacttgtga agggatcatt cttactgaat atgtctgttc tcttctatgt 540
actttctgac cctccagggc agctggatac agtggacctg aataagctaa acgagattga 600
aggcacccta aacaaagcca aagatgaaat gaaggtcagc gatcttgata ggaaagtgtc 660
tgacctggag aatgaagcca agaagcagga ggctgccatc atggactata accgagatat 720
cgaggagatc atgaaggaca ttcgcaatct ggaggacatc aggaagacct taccatctgg 780
ctgcttcaac accccgtcca ttgaaaagcc ctagtgtctt tagggctgga aggcagcatc 840
cctctgacag gggggcagtt gtgaggccac agagtgcctt gacacaaaga ttacattttt 900
cagacccccа ctcctctgct gctgtccatc actgtccttt tgaaccagga aaagtcacag 960
agtttaaaga gaagcaaatt aaacatcctg aatcgggaac aaagggtttt atctaataaa 1020
gtgtctcttc catcacgttg ctaccttacc cacacttccc tctgatttgc gtgaggacgt 1080
ggcatcctac ttacgtacgt ggcataacac atcgtgtgag cccatgtatg ctggggtaga 1140
gcaagtagcc ctcccctgtc tcatcgatcc agcagaacct cctcagtctc agtactcttg 1200
tttctataag gaaaagtttt gctactaaca gtagcattgt gatggccagt atatccagtc 1260
catggataaa gaaaatgcat ctgcatctcc tgcccctctt ccttctaagc aaaaggaaat 1320
aaacatcctg tgccaaaggt attggtcatt tagaatgtcg gtagccatcc atcagtgctt 1380
ttagctatta tgagtgtagg acactgagcc atccgtgggt caggatgcaa ttatttataa 1440
aagtccccag gtgaacatgg ctgaagattt ttctagtata ttaataattg actaggaaga 1500
tgaacttttt ttcagatctt tgggcagctg ataatttaaa tctggatggg cagcttgcac 1560
tcaccaatag accaaaagac atcttttgat attcttataa atggaactta cacagaagaa 1620
atagggatat gataaccact aaagttttgt tttcaaaatc aaactaattc ttacagcttt 1680
tttattagtt agtcttggaa ctagtgttaa gtatctggca gagaacagtt aatccctaag 1740
gtcttgacaa aacagaagaa aaacaagcct cctcgtccta gtcttttcta gcaaagggat 1800
aaaacttaga tggcagcttg tactgtcaga atcccgtgta tccatttgtt cttctgttgg 1860
agagatgaga catttgaccc ttagctccag ttttcttctg atgtttccat cttccagaat 1920
ccctcaaaaa acattgtttg ccaaatcctg gtggcaaata cttgcactca gtatttcaca 1980
cagctgccaa cgctatcgag ttcctgcact ttgtgattta aatccactct aaaccttccc 2040
tctaagtgta gagggaagac ccttacgtgg agtttcctag tgggcttctc aacttttgat 2100
cctcagctct gtggttttaa gaccacagtg tgacagttcc ctgccacaca ccccttcct 2160
cctaccaacc caccтttgag attcatatat agcctttaac actatgcaac tttgtacttt 2220
gcgtagcagg ggctggggtg gggggaaaga aacctattat catggacaca ctggtgctat 2280
taattatttc aaatttatat ttttgtgtga atgttttgtg ttttgtttat ccatgctata 2340
gaacaaggaa tttatgtaga tatacttagt cctatttcta gaatgacact ctgttcactt 2400
tgctcaattt ttcctcttca ctggcacaag tatctgaata cctccttccc tcccttctag 2460
agttctttgg attgtactcc aaagaattgt gccttgtgtt tgcagcatct ccattctcta 2520
aattaatata attgctttcc tccacaccca gccacgtaaa gaggtaactt gggtcctctt 2580
ccattgcagt cctgatgatc ctaacctgca gcacggtggt tttacaatgt tccagagcag 2640
gaacgccagg ttgacaagct atggtaggat taggaaagtt tgctgaagag gatctttgac 2700
gccacagtgg gactagccag gaatgaggga gaaatgccct ttttggcaat tgttggagct 2760
ggataggtaa gttttataag ggagtacatt ttgactgagc acttagggca tcaggaacag 2820
tgctacttac tggtgggtag actgggagag gtggtgtaac ttagttcttg atgatcccac 2880
ttcctgtttc catctgcttg ggatatacca gagtttacca caagtgtttt gacgatatac 2940
tcctgagctt tcactctgct ggcttctccc aggcctcttc tactatggca ggagatgtgg 3000
tgtgctgttg caaagttttc acgtcatcgt ttcctggcta gttcatttca ttaagtggct 3060
acatcctaac atatgcattg gtcaaggttg cagcaagagg actgaagatt gactgccaag 3120
ctagtttggg tgaagttcac tccagcaagt ctcaggccac aatggggtgg tttggtttgg 3180
tttcctttta actttctttt tgttatttgc ttttctcctc cacctgtgtg gtatattttt 3240
taagcagaat tttatttttt aaaataaaag gttctttaca agatgatacc ttaattacac 3300
tcccgcaaca cagccattat tttattgtct agctccagtt atctgtattt tatgtaatgt 3360
aattgacagg atggctgctg cagaatgctg gttgacacag ggattattat actgctattt 3420

ccggaggtga agtcggtgcc acttctgccc tggccccccaa gatcggcccc ctgggtctgt 180
ctccaaaaaa agttggtgat gacattgcca aggcaacggg tgactggaag ggcctgagga 240
ttacagtgaa actgaccatt cagaacagac aggcccagat tgaggtggtg ccttctgcct 300
ctgccctgat catcaaagcc ctcaaggaac caccaagaga cagaaagaaa cagaaaaaca 360
ttaaacacag tgggaatatc acttttgatg agattgtcaa cattgctcga cagatgcggc 420
accgatcctt agccagagaa ctctctggaa ccattaaaga gatcctgggg actgcccagt 480
cagtgggctg taatgttgat ggccgccatc ctcatgacat catcgatgac atcaacagtg 540
gtgctgtgga atgcccagcc agttaagcac aaaggaaaac atttcaataa aggatcattt 600
gacaactggt ga 612

<210> 17
<211> 1100
<212> DNA
<213> Homo sapiens

<400> 17
ctgcagtatt tagcatgccc cacccatctg caaggcattc tggatagtgt caaaacagcc 60
ggaaatcaag tccgtttatc tcaaacttta gcattttggg aataaatgat atttgctatg 120
ctggttaaat tagattttag ttaaatttcc tgctgaagct ctagtacgat aagcaacttg 180
acctaagtgt aaagttgaga tttccttcag gtttatatag cttgtgcgcc gcctgggtac 240
ctcggatgtg agggcgatct ggctgcgaca tctgtcaccc cattgatcgc cagggttgat 300
tcggctgatc tggctggcta ggcgggtgtc cccttcctcc ctcaccgctc catgtgcgtc 360
cctcccgaag ctgcgcgctc ggtcgaagag gacgaccatc cccgatagag gaggaccggt 420
cttcggtcaa gggtatacga gtagctgcgc tcccctgcta gaacctccaa acaagctctc 480
aaggtccatt tgtaggagaa cgtagggtag tcaagcttcc aagactccag acacatccaa 540
atgaggcgct gcatgtggca gtctgccttt cttttgaccc attacccatc taaggaccgg 600
tcttcggtca agggtatacg agtagctgcg ctcccctgct agaacctcca aacaagctct 660
caaggtccat ttgtaggaga acgtagggta gtcaagcttc caagactcca gacacatcca 720
aatgaggcgc tgcatgtggc agtctgcctt tcttttgacc cattacccat ctaagttaga 780
tgctttttta aatgtttttt aatttttaaa tttttaattt ttttcattat ttatttttta 840
ttttgagac ggatctcggc tcactgtaac ctccacctcc cgggttcaag cgattctcct 900
gactcagcct cccgagtagc tgggattaca ggcgcgccc accatgcccg gctaattttt 960
gtatttttag taaagacggg ttttcgccat gttgaccagg gtggtcttga actcctgacg 1020
tcaggtgatc ctcctgcctc gtgctgggat tgcaggcgtg agccaccgct cccggccctt 1080
agatgctttt taatcagcaa 1100

<210> 18
<211> 4747
<212> DNA
<213> Homo sapiens

<400> 18
taggaatgaa ccatccctgg gagtatgtgg aaacaacgga ggagctctga cttcccaact 60
gtcccattct atgggcgaag gaactgctcc tgacttcagt ggttaagggc agaattgaaa 120
ataattctgg aggaagataa gaatgattcc tgcgcgactg caccgggact acaaagggct 180
tgtcctgctg ggaatcctcc tggggactct gtgggagacc ggatgcaccc agatacgcta 240
ttcagttccg gaagagctgg agaaaggctc tagggtgggc gacatctcca gggacctggg 300
gctggagccc cgggagctcg cggagcgcgg agtccgcatc atccccagag gtaggacgca 360
gcttttcgcc ctgaatccgc gcagcggcag cttggtcacg gcgggcagga tagaccggga 420
ggagctctgt atgggggcca tcaagtgtca attaaatcta gacattctga tggaggataa 480
agtgaaaata tatggagtag aagtagaagt aagggacatt aacgacaatg cgccttactt 540
tcgtgaaagt gaattagaaa taaaaattag tgaaaatgca gccactgaga tgcggttccc 600
tctaccccac gcctgggatc cggatatcgg gaagaactct ctgcagagct acgagctcag 660
cccgaacact cacttctccc tcatcgtgca aaatggagcc gacggtagta agtaccccga 720
attggtgctg aaacgcgccc tggaccgcga agaaaaggct gctcaccacc tggtccttac 780
ggcctccgac gggggcgacc cggtgcgcac aggcaccgcg cgcatccgcg tgatggttct 840
ggatgcgaac gacaacgcac cagcgtttgc tcagcccgag taccgcgcga gcgttccgga 900
gaatctggcc ttgggcacgc agctgcttgt agtcaacgct accgaccctg acgaaggagt 960
caatgcggaa gtgaggtatt ccttccggta tgtggacgac aaggcggccc aagttttcaa 1020
actagattgt aattcaggga caatatcaac aataggggag ttggaccacg aggagtcagg 1080
attctaccag atggaagtgc aagcaatgga taatgcagga tattctgcgc gagccaaagt 1140
cctgatcact gttctggacg tgaacgacaa tgccccagaa gtggtcctca cctctctcgc 1200

<212> DNA
<213> Homo sapiens

<400> 19
gaattccgca gcccggagtg tggttagcag ctcggcaagc gctgcccagg tcctggggtg 60
gtggcagcca gcgggagcag gaaaggaagc atgttcccag gctgcccacg cctctgggtc 120
ctggtggtct tgggcaccag ctgggtagcc tgggggagcc aagggacaga agcggcacag 180
ctaaggcagt tctacgtggc tgctcagggc atcagttgga gctaccgacc tgagcccaca 240
aactcaagtt tgaatctttc tgtaacttcc tttaagaaaa ttgtctacag agagtatgaa 300
ccatatttta agaaagaaaa accacaatct accatttcag gacttcttgg gcctacttta 360
tatgctgaag tcggagacat cataaaagtt cactttaaaa ataaggcaga taagcccttg 420
agcatccatc ctcaaggaat taggtacagt aaattatcag aaggtgcttc ttaccttgac 480
cacacattcc ctgcggagaa gatggacgac gctgtggctc caggccgaga atacacctat 540
gaatggagta tcagtgagga cagtggaccc acccatgatg accctccatg cctcacacac 600
atctattact cccatgaaaa tctgatcgag gatttcaact cggggctgat tgggcccctg 660
cttatctgta aaaaagggac cctaactgag ggtgggacac agaagacgtt tgacaagcaa 720
atcgtgctac tatttgctgt gtttgatgaa agcaagagct ggagccagtc atcatcccta 780
atgtacacag tcaatggata tgtgaatggg acaatgccag atataacagt ttgtgcccat 840
gaccacatca gctggcatct gctgggaatg agctcggggc cagaattatt ctccattcat 900
ttcaacggcc aggtcctgga gcagaaccat cataaggtct cagccatcac ccttgtcagt 960
gctacatcca ctaccgcaaa tatgactgtg ggcccagagg gaaagtggat catatcttct 1020
ctcaccccaa aacatttgca agctgggatg caggcttaca ttgacattaa aaactgccca 1080
aagaaaacca ggaatcttaa gaaaataact cgtgagcaga ggcggcacat gaagaggtgg 1140
gaatacttca ttgctgcaga ggaagtcatt tgggactatg cacctgtaat accagcgaat 1200
atggacaaaa aatacaggtc tcagcatttg gataatttct caaaccaaat tggaaaacat 1260
tataagaaag ttatgtacac acagtacgaa gatgagtcct tcaccaaaca tacagtgaat 1320
cccaatatga aagaagatgg gattttgggt cctattatca gagcccaggt cagagacaca 1380
ctcaaaatcg tgttcaaaaa tatggccagc cgcccctata gcatttaccc tcatggagtg 1440
accttctcgc cttatgaaga tgaagtcaac tcttctttca cctcaggcag gaacaacacc 1500
atgatcagag cagttcaacc aggggaaacc tatacttata agtggaacat cttagagttt 1560
gatgaaccca cagaaaatga tgcccagtgc ttaacaagac catactacag tgacgtggac 1620
atcatgagag acatcgcctc tgggctaata ggactacttc taatctgtaa gagcagatcc 1680
ctggacaggc gaggaataca gagggcagca gacatcgaac agcaggctgt gtttgctgtg 1740
tttgatgaga acaaaagctg gtaccttgag gacaacatca acaagttttg tgaaaatcct 1800
gatgaggtga aacgtgatga ccccaagttt tatgaatcaa acatcatgag cactatcaat 1860
ggctatgtgc ctgagagcat aactactctt ggattctgct ttgatgacac tgtccagtgg 1920
cacttctgta gtgtggggac ccagaatgaa attttgacca tccacttcac tgggcactca 1980
ttcatctatg gaaagaggca tgaggacacc ttgaccctct tccccatgcg tggagaatct 2040
gtgacggtca caatggataa tgttggaact tggatgttaa cttccatgaa ttctagtcca 2100
agaagcaaaa agctgaggct gaaattcagg gatgttaaat gtatcccaga tgatgatgaa 2160
gactcatatg agatttttga acctccagaa tctacagtca tggctacacg gaaaatgcat 2220
gatcgtttag aacctgaaga tgaagagagt gatgctgact atgattacca gaacagactg 2280
gctgcagcat taggaattag gtcattccga aactcatcat tgaaccagga agaagaagag 2340
ttcaatctta ctgccctagc tctggagaat ggcactgaat tcgtttcttc gaacacagat 2400
ataattgttg gttcaaatta ttcttcccca agtaatatta gtaagttcac tgtcaataac 2460
cttgcagaac ctcagaaagc cccttctcac caacaagcca ccacagctgg ttccccactg 2520
agacacctca ttggcaagaa ctcagttctc aattcttcca cagcagagca ttccagccca 2580
tattctgaag accctataga ggatcctcta cagccagatg tcacagggat acgtctactt 2640
tcacttggtg ctggagaatt cagaagtcaa gaacatgcta agcgtaaggg acccaaggta 2700
gaaagagatc aagcagcaaa gcacaggttc tcctggatga aattactagc acataaagtt 2760
gggagacacc taagccaaga cactggttct ccttccggaa tgaggccctg ggaggacctt 2820
cctagccaag acactggttc tccttccaga atgaggccct gggaggaccc tcctagtgat 2880
ctgttactct taaaacaaag taactcatct aagattttgg ttgggagatg gcatttggct 2940
tctgagaaag gtagctatga aataatccaa gatactgatg aagacacagc tgttaacaat 3000
tggctgatca gcccccagaa tgcctcacgt gcttggggag aaagcacccc tcttgccaac 3060
aagcctggaa agcagagtgg ccacccaaag tttcctagag ttagacataa atctctacaa 3120
gtaagacagg atggaggaaa gagtagactg aagaaaagcc agtttctcat taagacacga 3180
aaaaagaaaa aagagaagca cacacaccat gctcctttat ctccgaggac ctttcaccct 3240
ctaagaagtg aagcctacaa cacattttca gaaagaagac ttaagcattc gttggtgctt 3300
cataaatcca atgaaacatc tcttcccaca gacctcaatc agacattgcc ctctatggat 3360
tttggctgga tagcctcact tcctgaccat aatcagaatt cctcaaatga cactggtcag 3420
gcaagctgtc ctccaggtct ttatcagaca gtgcccccag aggaacacta tcaaacattc 3480
cccattcaag accctgatca aatgcactct acttcagacc ccagtcacag atcctcttct 3540
ccagagctca gtgaaatgct tgagtatgac cgaagtcaca agtccttccc cacagatata 3600

ccgccgcctt ctactccgcc gcgggggtcg cagcggctgc cgcgccgtcc tcgagtttcc 60
agcgtgagga ggaggctgag ggcggagagg cgcatcgtgt tcgaggcgga gaccgagggg 120
gagccccgcg cgcggcgtcg ctcattgcta tggacagtgc tatcaccctg tggcagttcc 180
ttcttcagct cctgcagaag cctcagaaca agcacatgat ctgttggacc tctaatgatg 240
ggcagtttaa gcttttgcag gcagaagagg tggctcgtct ctgggggatt cgcaagaaca 300
agcctaacat gaattatgac aaactcagcc gagccctcag atactattat gtaaagaata 360
tcatcaaaaa agtgaatggt cagaagtttg tgtacaagtt tgtctcttat ccagagattt 420
tgaacatgga tccaatgaca gtgggcagga ttgagggtga ctgtgaaagt ttaaacttca 480
gtgaagtcag cagcagttcc aaagatgtgg agaatggagg gaaagataaa ccacctcagc 540
ctggtgccaa gacctctagc cgcaatgact acatacactc tggcttatat tcttcattta 600
ctctcaactc tttgaactcc tccaatgtaa agcttttcaa attgataaag actgagaatc 660
cagccgagaa actggcagag aaaaaatctc ctcaggagcc cacaccatct gtcatcaaat 720
ttgtcacgac accttccaaa aagccaccag ttgaacctgt tgctgccacc atttcaattg 780
gcccaagtat ttctccatct tcagaagaaa ctatccaagc tttggagaca ttggtttccc 840
caaaactgcc ttccctggaa gccccaacct ctgcctctaa cgtaatgact gcttttgcca 900
ccacaccacc catttcgtcc ataccccctt tgcaggaacc tcccagaaca ccttcaccac 960
cactgagttc tcacccagac atcgacacag acattgattc agtggcttct cagccaatgg 1020
aacttccaga gaatttgtct ctggagccta aagaccagga ttcagtcttg ctagaaaagg 1080
acaaagtaaa taattcatca agatccaaga aacccaaagg gttaggactg gcacccaccc 1140
ttgtgatcac gagcagtgat ccaagcccac tgggaatact gagcccatct ctccctacag 1200
cttctcttac accagcattt ttttcacaga cacccatcat actgactcca agccccttgc 1260
tctccagtat ccacttctgg agtactctca gtcctgttgc tcccctaagt ccagccagac 1320
tgcaaggtgc taacacactt ttccagtttc cttctgtact gaacagtcat gggccattca 1380
ctctgtctgg gctggatgga ccttccaccc ctggcccatt ttccccagac ctacagaaga 1440
cataacctat gcacttgtgg aatgagagaa ccgaggaacg aagaaacaga cattcaacat 1500
gattgcattt gaagtgagca attgatagtt ctacaatgct gataatagac tattgtgatt 1560
tttgccattc cccattgaaa acatcttttt aggattctct ttgaatagga ctcaagttgg 1620
actatatgta taaaaatgcc ttaattggag tctaaactcc acctccctct gtcttttcct 1680
tttctttttc tttccttcct tcctttttct ttctccttta aaaatatttt gagctttgtg 1740
ctgaagaagt ttttggtggg ctttagtgac tgtgctttgc aaaagcaatt aagaacaaag 1800
ttactccttc tggctattgg gacccttggg ccaggaaaaa ttatgcttag aatctattat 1860
ttaaagaagt atttgtgaaa tgaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1920
aaaaaaaaaa aaa 1933

<210> 21
<211> 2215
<212> DNA
<213> Homo sapiens

<400> 21
actcagtaga ccgccactgg ctgtgcacgt tatggggttt ccacctaggg ctcggcctga 60
ggcttgtaac actccgtttt cccccgagtc acaggggcag tcttgccccc cgcagctggg 120
tcgcggtgtc tctcaaaggt ccccctctac aggggcttcg tgaggcccgg gcccacaggg 180
cgctcggtcc cggaagtgac gtctcccaga ggggccggaa gtggcagtgg agggagggaa 240
gatggcggag gtgggggaga taatcgaggg ctgccgccta cccgtgctgc ggcggaacca 300
ggacaacgaa gatgagtggc ccctggccga gatcctgagc gtgaaggaca tcagtggccg 360
gaagcttttc tacgtccatt acattgactt caacaaacgt ctggatgaat gggtgacgca 420
tgagcggctg gacctaaaga agatccagtt ccccaagaaa gaggccaaga cccccactaa 480
gaacggactt cctgggtccc gtcctggctc tccagagaga gaggtgccgg cctcggcgca 540
ggccagcggg aagaccttgc caatcccggt ccagatcaca ctccgcttca acctgcccaa 600
ggagcgggag gccattcccg gtggcgagcc tgaccagccg ctctcctcca gctcctgcct 660
gcagcccaac caccgctcaa cgaaacggaa ggtggaggtg gtttcaccag caactccagt 720
gcccagcgag acagccccgg cctcggtttt tccccagaat ggagccgccc gtagggcagt 780
ggcagcccag ccaggacgga agcgaaaatc gaattgtttg ggcactgatg aggactccca 840
ggacagctct gatggaatac cgtcagcacc acgcatgact ggcagcctgg tgtctgatcg 900
aagccacgac gacatcgtca cccggatgaa gaacattgag tgcattgagc tgggccggca 960
ccgcctcaag ccgtggtact tctccccgta cccacaggaa ctcaccacat tgcctgtcct 1020
ctacctgtgc gagttctgcc tcaagtacgg ccgtagtctc aagtgtcttc agcgtcattt 1080
gaccaagtgt gacctacgac atcctccagg caatgagatt taccgcaagg gcaccatctc 1140
cttctttgag attgatggac gtaagaacaa gagttattcc cagaacctgt gtcttttggc 1200
caagtgtttc cttgaccata agacactgta ctatgacaca gacccttcc tcttctacgt 1260
catgacagag tatgactgta agggcttcca catcgtgggc tacttctcca aggagaaaga 1320
atcaacggaa gactacaatg tggcctgcat cctaaccctg cctccctacc agcgccgggg 1380

ggtctctggg gaacacagcc agacatcctg atggggaggc agagccagga agctaagcca 1320
gggcccagct gcgtccaacc cagcccccca cctcaggtcc ctgaccccag ctcgatgccc 1380
catcaccgcc tctccctggc tcccaagggt gcaggtgggc gcaaggcccg gcccccatca 1440
catgttccct tggcctcaga gctgcccctg ctctcccacc acagccaccc agaggcaccc 1500
catgaagctt ttttctcgtt cactcccaaa cccaagtgtc caaagctcca gtcctaggag 1560
aacagtccct gggtcagcag ccaggaggcg gtccataaga atggggacag tgggctctgc 1620
cagggctgcc gcacctgtcc agaacaacat gttctgttcc tcctcctcat gcatttccag 1680
ccttg 1685

<210> 24
<211> 6428
<212> DNA
<213> Homo sapiens

<400> 24
gctagtggaa gttactgccg cgccaccgag tccggaccgg agactttggg gcctaactag 60
tgaatggtag tgtctagaaa gggtatgtcc cttcaagaga gaggtgccaa tgtccaaccg 120
gcctaataac aatccagggg ggtcactgcg acgttcacag aggaacactg ccggggccca 180
accacaagac gactcaatag gaggaagaag ctgcagttca tcatctgctg tgatagttcc 240
acaaccagag gatccagaca gagccaatac ttcagaaaga caaaaaacgg ggcaggtgcc 300
taagaaagac aattctcgag gagtgaagcg cagtgctagt ccagactaca acaggaccaa 360
ttctcctagc tctgcaaaaa aaccaaaagc acttcagcat actgaatctc cctcagaaac 420
aaataagcca catagtaagt caaagaagag acatttagac caggagcaac aactgaaatc 480
tgcacaatca ccatcaacaa gcaaggctca taccaggaag agtggggcca ctggcggttc 540
acggagtcag aaaagaaaaa ggacagagag ttcttgtgta aagagtggct ccgggtctga 600
atcaactggt gcagaagaga gatctgcgaa acctaccaag ctggcttcaa aatcagccac 660
ctcagccaaa gctgggtgta gcaccatcac tgattcttct tctgctgcct ctacttcctc 720
ctcgtcttct gctgtagcct cggcctcctc cactgtacca ccaggtgcca gagtgaaaca 780
aggaaaagat cagaacaagg ccaggcgttc ccgttcagcg tccagtccca gccccagaag 840
aagtagcagg gaaaaggaac agagtaaaac tggtggctct tcaaaatttg attgggctgc 900
tcgtttcagc cctaaagtta gccttcctaa aacaaaactg tctcttccag ggtcttctaa 960
gtcagagaca tcaaaacctg gaccttctgg attacaggcc aaattagcaa gtttaagaaa 1020
atctacgaag aaacgcagtg agtctccacc tgctgagctc cccagtttga ggcggagcac 1080
acgccaaaag accacgggct cctgtgctag taccagtcgg cgaggctctg gcctgggcaa 1140
aagaggagca gctgaagctc gtcgacagga gaaaatggca gaccctgaaa gcaaccagga 1200
ggcagtaaat tcttcagctg ctcggacaga tgaagctccc caaggagctg caggggctgt 1260
tggcatgacc acctctgggg agagtgaatc agatgattcc gagatgggac gtttgcaagc 1320
tttgttagag gcaaggggtc ttccccctca cctatttggt cctcttggtc ctcggatgtc 1380
acagcttttc catagaacaa ttggaagtgg agctagttct aaggcccagc agctactaca 1440
aggattgcaa gccagtgatg aaagtcaaca gcttcaggca gttattgaga tgtgtcagtt 1500
actggtcatg ggaaatgagg agacactggg agggtttcct gtcaagagtg ttgttccagc 1560
tttgattacg ttacttcaga tggagcacaa ttttgatatt atgaaccatg cttgtcgagc 1620
cttaacatac atgatggaag cacttcctcg atcttctgct gttgtagtag atgctattcc 1680
tgtctttta gaaaagctgc aagttattca gtgtattgat gtggcagagc aggccttgac 1740
tgccttggag atgttgtcac ggagacatag taaagccatt ctacaggcgg gtggtttggc 1800
agactgcttg ctgtacctag aattcttcag cataaatgcc caaagaaatg cattagcaat 1860
tgcagctaat tgctgccaga gtatcacgcc agatgaattt cattttgtgg cagattcact 1920
cccattgcta acccaaaggc taacacatca ggataaaaag tcagtagaaa gcacttgcct 1980
ttgttttgca cgcctagtgg acaacttcca gcatgaggag aatttactcc agcaggttgc 2040
ttccaaagat ctgcttacaa atgttcaaca gctgttggta gtgactccac ccattttaag 2100
ttctgggatg tttataatgg tggttcgcat gttttctctg atgtgttcca actgtccaac 2160
tttagctgtt caacttatga aacaaaacat tgcagaaacg cttcactttc tcctgtgtgg 2220
tgcctccaat ggaagttgtc aggaacagat tgatcttgtt ccacgaagcc ctcaagagtt 2280
gtatgaactg acatctctga tttgtgaact tatgccatgt ttaccaaaag aaggcatttt 2340
tgcagttgat accatgttga agaagggaaa tgcacagaac acagatggtg cgatatggca 2400
gtggcgtgat gatcggggcc tctggcatcc atataacagg attgacagcc ggatcattga 2460
gcaaatcaat gaggacacgg gaacagcacg tgccattcag agaaaaccta acccgttagc 2520
caatagtaac actagtggat attcagagtc aaagaaggat gatgctcgag cacagcttat 2580
gaaagaggat ccggaactgg ctaagtcttt tattaagaca ttatttggtg ttctttatga 2640
agtgtatagt tcctcagcag gacctgcggt cagacataag tgccttagag caattcttag 2700
gataatttat tttgcggatg ctgaacttct gaaggatgtt ctgaaaaatc atgctgtttc 2760
aagtcacatt gcttccatgc tgtcaagcca agacctgaag atagtagtgg gagcacttca 2820
gatggcagaa attttaatgc agaagttacc tgatattttt agtgtttact tcagaagaga 2880

<212> DNA
<213> Homo sapiens

<400> 25
gcacgaggcc agccttacgg gcccgaaccc tcgtgtgaag ggtgcagtac ctaagccgga 60
gcggggtaga ggcgggccgg cacccccttc tgacctccag tgccgccggc ctcaagatca 120
gacatggccc agaacttgaa ggacttggcg ggacggctgc ccgccgggcc ccggggcatg 180
ggcacggccc tgaagctgtt gctgggggcc ggcgccgtgg cctacggtgt gcgcgaatct 240
gtgttcaccg tggaaggcgg gcacagagcc atcttcttca atcggatcgg tggagtgcag 300
caggacacta tcctggccga gggccttcac ttcaggatcc cttggttcca gtaccccatt 360
atctatgaca ttcgggccag acctcgaaaa atctcctccc ctacaggctc caaagaccta 420
cagatggtga atatctccct gcgagtgttg tctcgaccca atgctcagga gcttcctagc 480
atgtaccagc gcctagggct ggactacgag gaacgagtgt tgccgtccat tgttaatgag 540
gtgctcaaga gtgtggtggc caagttcaat gcctcacagc tgatcaccca gcgggcccag 600
gtatccctgt tgatccgccg ggagctgaca gagagggcca aggacttcag cctcatcctg 660
gatgatgtgg ccatcacaga gctgagcttt agccgagagt acacagctgc tgtagaagcc 720
aaacaagtgg cccagcagga ggcccagcgg gcccaattct tggtagaaaa agcaaagcag 780
gaacagcggc agaaaattgt gcaggccgag ggtgaggccg aggctgccaa gatgcttgga 840
gaagcactga gcaagaaccc tggctacatc aaacttcgca agattcgagc agcccagaat 900
atctccaaga cgatcgccac atcacagaat cgtatctatc tcacagctga caaccttgtg 960
ctgaacctac aggatgaaag tttcaccagg ggaagtgaca gcctcatcaa gggtaagaaa 1020
tgagcctagt caccaagaac tccaccccca gaggaagtgg atctgcttct ccagtttttg 1080
aggagccagc caggggtcca gcacagccct accccgcccc agtatcatgc gatggtcccc 1140
cacaccggtt ccctgaaccc ctcttggatt aaggaagact gaagactagc ccctttctg 1200
gggaattact ttcctcctcc ctgtgttaac tggggctgtt ggggacagtg cgtgatttct 1260
cagtgatttc ctacagtgtt gttccctccc tcaaggctgg gaggagataa acaccaaccc 1320
aggaattctc aataaatttt tattacttaa cctgaaaaaa aaaaaaaaaa 1370

<210> 26
<211> 642
<212> DNA
<213> Homo sapiens

<400> 26
ccgcgcgtca cgtgacccca gcgcctactt gggctgagga gccgccgcgt cccctcgccg 60
agtcccctcg ccagattccc tccgtcgccg ccaagatgat gtgcggggcg ccctccgcca 120
cgcagccggc caccgccgag acccagcaca tcgccgacca ggtgaggtcc cagcttgaag 180
agaaagaaaa caagaagttc cctgtgttta aggccgtgtc attcaagagc caggtggtcg 240
cggggacaaa ctacttcatc aaggtgcacg tcggcgacga ggacttcgta cacctgcgag 300
tgttccaatc tctccctcat gaaaacaagc ccttgacctt atctaactac cagaccaaca 360
aagccaagca tgatgagctg acctatttct gatcctgact ttggacaagg cccttcagcc 420
agaagactga caaagtcatc ctccgtctac cagagcgtgc acttgtgatc ctaaaataag 480
cttcatctcc gctgtgccct tggggtggaa ggggcaggat tctgcagctg cttttgcatt 540
tctcttccta aatttcattg tgttgatttc tttccttccc aataggtgat cttaattact 600
ttcagaatat tttcaaaata gatatatttt taaaatcctt ac 642